



Pro Gradu - tutkielma  
Ilmakehtätieteiden maisteriohjelma  
Meteorologia

## **Pintasääsuureiden vertailu eri analyysijärjestelmissä**

Oskari Rockas  
3.12.2019

Ohjaajat: Jussi Ylhäisi  
Heikki Järvinen  
-  
Tarkastajat: Heikki Järvinen  
Jouni Räisänen

HELSINGIN YLIOPISTO  
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA

PL 64 (Gustaf Hällströmin katu 2a)  
00014 Helsingin yliopisto

Tiedekunta — Fakultet — Faculty		Koulutusohjelma — Utbildningsprogram — Degree programme	
Matemaattis-luonnontieteellinen		Ilmakehätieteiden maisteriohjelma	
Tekijä — Författare — Author			
Oskari Rockas			
Työn nimi — Arbetets titel — Title			
Pintasääsuureiden vertailu eri analyysijärjestelmissä			
Oppiaine — Läroämne — Subject			
Meteorologia			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu -tutkielma		Joulukuu 2019	46 s.
Tiivistelmä — Referat — Abstract			
<p>Meteorologisten suureiden, kuten lämpötilan, kaksi- tai kolmiulotteisia hilakenttiä tuotetaan analyysijärjestelmien avulla. Näitä kenttiä käytetään esimerkiksi päivittämään lähituntien sääennusteita ja tarkan resoluution vuoksi niiden avulla voidaan myös tutkia mm. mesoskaalan ilmiöitä. Ilmatieteen laitoksella on käytössään useita erilaisia analyysi- ja nowcast-malleja, joiden osuvuutta tässä tutkielmassa selvitetään. Tarkoituksena on luoda kokonaiskuva kolmen eri järjestelmän pinta-analyysien vahvuuksista ja heikkouksista Suomen alueella.</p> <p>Analyysivertailuun valikoitui malleja, joista kahden kehitykseen tai optimointiin on osallistuttu Ilmatieteen laitoksella (LAPS-analyysimalli ja pohjoismaiseen yhteistyöhön perustuva MEPS-nowcast) ja yhtä kehitetään Norjassa (ns. METNO-analyysi). METNO:n lämpötila-analyysissa hyödynnetään perinteisten havaintolähteiden lisäksi kansalaisten omistamien Netatmo-asemien keräämää havaintotietoa, minkä vuoksi se valikoitui mukaan verifiointiin. 434 Suomen alueella sijaitsevan Netatmo-aseman havaintoja päädyttiin myös osin käyttämään tutkielman verrokkiaineistona 197 suomalaisen SYNOP-aseman lisäksi, sillä näin voitiin hyödyntää malleille riippumatonta havaintoaineistoa METNO:n lämpötila-analyysia lukuun ottamatta.</p> <p>Suureista lämpötilan ja merenpintapaineen analyysit ovat laadukkaimpia perustuen tässä tutkimuksessa tehtyyn lineaariseen regressioanalyysiin, kun taas kosteuden ja tuulen kohdalla hajontaa mallin ja havaintojen välillä on runsaammin. Paineanalyysien välillä ei ole juurikaan eroja, kun taas selkein ero kyvykkyydessä on huomattavissa tuulianalyyseissa, joista LAPS-analyysien verifiointitulokset ovat parhaimmat. Kahdesta LAPS-versiosta MEPS-säämallia taustakenttänään käyttävä LAPS-analyysi on tässä tutkimuksessa myös laadukkein kosteuden tapauksessa sekä riippuvan että riippumattoman havaintovertailun perusteella. Maantieteellisesti haastavimmiksi alueiksi analyysille on tutkimuksessa havaittu Lappi sekä LAPS-järjestelmän tapauksessa rannikkoalueet.</p> <p>Norjalaista lämpötila-analyysia voidaan pitää etenkin SYNOP-vertailun pohjalta analyyseista parhaimpana. Ilman laatuvarmistettua ja riippumatonta asemavertailua on kuitenkin vaikea määrittää, mikä on Netatmo-asemien tuoma lisäarvo analyysissa. Tulevaisuudessa laadukkaan ja riippumattoman havaintoaineiston käyttö täydentäisi hyvin tätä tutkimusta ja se voitaisiin suorittaa esimerkiksi pidättämällä pieni SYNOP-asemajoukko analyyseista.</p>			
Avainsanat — Nyckelord — Keywords			
analyysimalli, data-assimilaatio, nowcast, verifiointi, LAPS, MEPS, Netatmo			
Säilytyspaikka — Förvaringsställe — Where deposited			
Kumpulan tiedekirjasto			
Muita tietoja — Övriga uppgifter — Additional information			

# Sisältö

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Data-assimilaatio säännestamisessä</b>	<b>3</b>
2.1	Data-assimiloinnin periaatteet . . . . .	3
2.2	Empiirinen analyysi . . . . .	5
2.2.1	Perättäinen korjausmetodi . . . . .	5
2.3	Tilastollinen analyysi . . . . .	6
2.3.1	Optimaalinen interpolointi . . . . .	7
2.3.2	Kalman-suodatin . . . . .	8
2.4	Parvimenetelmät . . . . .	9
<b>3</b>	<b>Aineisto</b>	<b>12</b>
3.1	Havaintoverkostot . . . . .	13
3.2	Analyysi- ja nowcastmallit . . . . .	14
3.2.1	Local Analysis and Prediction System (LAPS) . . . . .	14
3.2.2	MET Nordic Analysis (METNO) . . . . .	15
3.2.3	MEPS-nowcast (MNWC) . . . . .	16
<b>4</b>	<b>Menetelmät</b>	<b>18</b>
4.1	Tilastollinen lähestymistapa . . . . .	18
4.2	Kastepisteen ja merenpintapaineen laskentakaavat . . . . .	20
<b>5</b>	<b>Tulokset</b>	<b>21</b>
5.1	Pintasuureiden vertailu riippuvassa havaintoverkostossa . . . . .	21
5.1.1	Lämpötila . . . . .	22
5.1.2	Merenpintapaine . . . . .	23
5.1.3	Suhteellinen kosteus ja kastepiste . . . . .	25
5.1.4	Tuulen nopeus . . . . .	28
5.2	Pintasuureiden vertailu Netatmo-asemien avulla . . . . .	30
<b>6</b>	<b>Pohdinta</b>	<b>33</b>
6.1	Johtopäätökset data-analyysistä . . . . .	33

---

6.1.1	Kokonaiskuva suureiden osalta ja Netatmo-asemien tuoma lisäarvo	33
6.1.2	Maantieteelliset ongelma-alueet ja lisätarkastelut . . . . .	35
6.2	Vertailu aiempaan tutkimustietoon . . . . .	37
6.3	Tutkimuksen arviointi ja jatkoehdotukset . . . . .	38
<b>7</b>	<b>Yhteenveto</b>	<b>40</b>
	<b>Kirjallisuutta</b>	<b>42</b>
	<b>Liitteet</b>	<b>47</b>



# 1. Johdanto

Ilmakehä on kaaottinen systeemi, joka on jatkuvassa vuorovaikutuksessa maanpinnan sekä merialueiden kanssa. Ymmärryksemme ilmakehän vallitsevasta tilasta ja käyttäytymisestä tulevaisuudessa perustuu karkeasti sanottuna kahteen tietolähteeseen: havaintoihin, jotka kuvaavat ilmakehän vallitsevan tilan sekä malleihin, jotka tuottavat arvauksen ilmakehän tulevasta käyttäytymisestä (Lahoz ja Schneider, 2014). Johtuen esimerkiksi pintasäähavaintoverkoston rajallisesta kattavuudesta (virallisia sääasemia maapallolla on n. 11000 kpl; National Research Council, 2009) ei havaintojen avulla pystytä täydellisesti kuvaamaan ilmakehän tilaa. Jos jo mallin arvaus ilmakehän vallitsevasta alkutilasta on kovin virheellinen, heijastuu tämä ongelmina mallin tuottamassa ennusteessa; tämän vuoksi ilmakehän analyysien parantaminen erilaisten data-assimilaatiomenetelmien avulla on noussut tärkeäksi tutkimusaiheeksi meteorologian saralla (Daley, 1993; Talagrand, 1997).

Analyysijärjestelmäksi- tai malliksi kutsutaan säänennustamisen tapauksessa mallia, joka tuottaa nk. data-assimilaation avulla meteorologisten parametrien hilamuotoisia kenttiä vallitsevalle ajanhetkelle (Gregow, 2018). Analyysimallien avulla on mahdollista tuottaa hienohilaisia kenttiä esimerkiksi mesoskaalan tai sitä pienempien ilmiöiden tutkimiseen (Koskinen et al., 2011, Häggmark et al., 2000), ja säämallien lisäksi niitä voidaan hyödyntää mm. maanpintamallien alustamisessa (Hiemstra et al., 2006). Tällöin säätilan analyyseista voi olla yleishyödyllistä apua niin tulvimisen (Jasper et al., 2002) kuin vaikka lumimäärien kehittymisen arvionnissa (Liston ja Elder, 2006).

Nykyään yksi analyysijärjestelmien käyttökohteista on olla tärkeä apuväline hyvin lyhyen aikavälin säänennustamisessa. Tuoreimman havaintotiedon avulla tuotetuilla analyysilla voidaan korjata ennusteen alkuhetken virheitä; lisäksi analyysimenetelmiä voidaan hyödyntää ns. nowcasting-malleissa, jotka tuottavat analyysit lähtökohtanaan lyhyitä, 0-24 tunnin pituisia ennusteita. Nowcasting-malleilla on mahdollista tuottaa pitkän aikavälin ennustemalleja spatiaalisesti tarkempia ennusteita, jolloin esimerkiksi useimmille säämalleille hankalat konvektiiviset prosessit ovat paremmin mallinnettavissa (Gregow, 2018). Nowcasting-mallien kehityksessä juuri syvän konvektion mallintaminen on ollut usein edistävänä tekijänä (Wilson et al., 2004; Dance et al., 2010; Ruzans-

ki et al., 2011), mutta myös laajempiin käyttötarkoituksiin kehitetyt järjestelmät ovat yleistyneet (esimerkiksi Haiden et al., 2011).

Sääpalveluja tuottavat kansalliset keskuskeskukset usein kehittävät ja hyödyntävät työssään montaa eri analyysi- ja nowcast-mallia samanaikaisesti. Tämän tutkielman tarkoituksena on vertailla ja verifoida kolmea eri Ilmatieteen laitoksella hyödynnettävää järjestelmää, joista kahden kehittämiseen tai käyttöönottoon on osallistuttu suoraan Ilmatieteen laitoksella (analyysijärjestelmä LAPS sekä pohjoismaisen yhteistyön MEPS:n nowcast-malli) ja joista kolmas on kehitetty Norjan ilmatieteen laitoksella. Näistä järjestelmistä LAPS on ollut laajemminkin maailmalla käytössä, ja siksi sen analyysien laadusta on olemassa aiempaa tutkimusta. Esimerkiksi Hiemstra et al. (2006) tutkimuksessa LAPS-analyysin kyvykkyyttä lämpötilan, tuulen, kosteuden sekä sademäärän osalta verrattiin itsenäisiin havaintoihin kahden vuoden ajan Yhdysvaltojen keskilänessä, kun taas Vitali et al. (2008) tarkastelivat raportissaan miten LAPS:n tuuli- ja lämpötila-analyysit suhteutuivat kahden Pohjois-Italiassa sijaitsevan itsenäisen havaintoaseman havaintotietoon vuodelta 1999 vertaamalla lisäksi LAPS-analyysseja toiseen analyysiin. Lineaariseen regressioanalyysiin pohjaten sekä Hiemstra et al. (2006) että Vitali et al. (2008) havaitsivat lämpötila-analyysin olevan ainakin LAPS-järjestelmän tapauksessa tuulianalyysia laadukkaampi.

Norjalainen analyysijärjestelmä on kiinnostava tutkimuskohde siinä mielessä, että siihen syötetään ns. kolmannen osapuolen havaintoaineistoa eli kansalaisten omistamien Netatmo-asemien havaintotietoa. Esimerkiksi Chapman et al. (2017) ovat tutkineet Netatmo-havaintojen hyödyntämistä Lontoon lämpösaarekkeen voimakkuuden arvioinnissa ja totesivat Netatmo-asemien hyödyntämisessä olevan paljon mahdollisuuksia huomioiden aineiston laatuvarmistuksen tason. Tässä tutkielmassa pyritään selvittämään, vaikuttaako kansalaishavainnoista olevan lisäarvoa analyysissa lähtien siitä hypoteesista liikkeelle, että Netatmo-asetat kasvattavat norjalaisen analyysin kyvykkyyttä muihin nähden. Ylipäättään tutkielmassa pyritään antamaan kokonaiskuva siitä, miten tutkittavat järjestelmät onnistuvat neljän eri pintasuureen (lämpötila, merenpintapaine, suhteellinen kosteus ja tuulen nopeus) analysoinnissa Suomen alueella vertaamalla niitä 197 suomalaisen säähavaintoaseman ja osin 494 Netatmo-aseman havaintoihin sekä selvittämällä, onko aineistossa havaittavissa maantieteellisiä ongelma-alueita.

Luvussa 2 esitellään sekä data-assimiloinnin peruseriaatteet että tutkittavissa järjestelmissä käytettävien data-assimilaatiomenetelmien taustateoriat. Luvussa 3 käsitellään tarkemmin verifioitavat järjestelmät sekä niihin syötettävät havaintotiedot. Luvussa 4 käydään läpi tilastolliset menetelmät data-analyysin taustalla ja luvussa 5 taas esitellään verifiointin tulokset pintasuure kerrallaan. Luvut 6 ja 7 on omistettu tulosten pohdinnalle sekä yhteenvedolle.

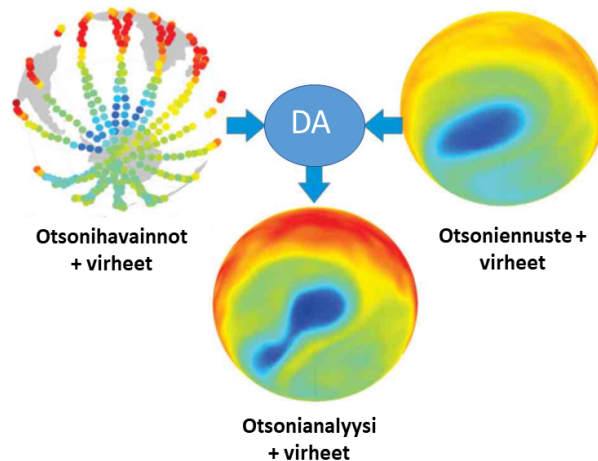
## 2. Data-assimilaatio säänennustamisessa

Tässä luvussa käydään läpi data-assimilaatio säänennustamisessa sekä tutkielman kannalta tarpeelliset data-assimilaatiomenetelmät. Aluksi esitellään data-assimilaation peruseriaatteita (luku 2.1), minkä jälkeen empiirisestä analyysistä (luku 2.2) siirrytään tilastollisen analyysin esittelyyn (luku 2.3). Empiirisestä analyysimenetelmästä esimerkkinä on perättäinen korjausmetodi (luku 2.2.1) ja tilastollisesta analyysistä taas optimaalinen interpolointi (luku 2.3.1) sekä Kalman-suodatin (luku 2.3.2). Lopuksi tutustutaan viime vuosikymmeninä yleistyneisiin parvimenetelmiin (luku 2.4).

### 2.1 Data-assimiloinnin periaatteet

Nykypäivänä ilmakehästä on saatavilla monipuolisesti havaintotietoa: ilmakehää havainnoidaan niin perinteisten pistehavaintojen (mm. SYNOP-asemaverkosto) kuin laajemmin esimerkiksi satellittien avulla. Kaikille havaintolähteille on yhteistä kuitenkin se, että niiden käytössä on rajoitteita (Lahoz ja Schneider, 2014); mittaukset ja havaintolaitteet itsessään sisältävät erisuuruisia virheitä (systemaattisia tai satunnaisia, Lahoz, 2010), ja niiden edustavuus ajan tai paikan suhteen ei ole itsessään riittävällä tasolla säänennustamisen tarpeita ajatellen (Kalnay, 2003). Lisäksi havaintoihin liittyy mittausvirheistä riippumattomia *edustavuusvirheitä* (Janjić et al., 2018); ongelmaksi voi muodostua esimerkiksi se, jos havaintoihin vaikuttaa voimakkaasti säämallin tai analyysin hilakokoa pienempiä ilmiöitä. Data-assimilaatiomenetelmien avulla pyritään levittämään havaintojen sisältämä tieto alueille, joissa havaintotietoja ei ole käytettävissä yhdistämällä havaintotieto virheineen jonkinlaiseen mallin arvaukseen virheineen (kuva 2.1, Lahoz ja Schneider, 2014).

Säänennustamisen tapauksessa data-assimilaatio alkaa mallin arvauksesta ilmakehän nykyisestä tilasta, jota kutsutaan *taustakentäksi*. Taustakenttä on tarpeellinen, sillä kun levitetään havaintojen sisältämää informaatiota ympäristöön eli interpoloidaan havaintoja säämallin alkutilan määrittämiseksi, on havaintojen keskimääräinen lukumäärä kahta kertaluokkaa pienempi säämallin vapausasteisiin verrattuna (Kalnay, 2003). Taustakenttänä käytetään yleensä valitun säämallin viimeisintä vallitsevalle ajanhetkelle



**Kuva 2.1:** Havaintotiedon, mallitiedon ja niiden virheiden yhdistäminen analyysiksi data-assimilaation (DA) avulla kuvattuna stratosfäärin otsonin avulla. Sininen väri kuvaa suhteellisesti matalia, punainen korkeita pitoisuuksia. Lähde: Lahoz ja Schneider, 2014

tuotettua ennustetta; tämän vuoksi taustakenttä on tavallisesti jokin useita kertoja vuorokaudessa päivittyvä lyhyen aikavälin säämalli (Lahoz ja Schneider, 2014). Taustakenttä on ensimmäinen arvaus ilmakehän nykyisestä tilasta, ja sitä korjataan viimeisimpien havaintojen avulla vastaamaan paremmin ilmakehän todellista tilaa. Tämä prosessi on kuvattuna seuraavassa yhtälössä (Kalnay, 2003):

$$(2.1) \quad x^a = x^b + W[y^o - H(x^b)]$$

jossa  $x$  on kolmiulotteinen ilmakehän tilamuuttuja, yläindeksi  $a$  kuvaa data-assimilaation lopputuotetta analyysia ja  $b$  taas taustakenttää.  $y^o$  on käytetyt havainnot ja  $H$  havainto-operaattori, joka interpoloi sekä prosessoi taustakentän malliavaruudesta havaintoavaruuteen. Tekijä  $H(x^b)$  on havainnon mallivastine havaintoavaruudessa, ja kokonaisuudessaan  $y^o - H(x^b)$  kuvaa *innovaatiota* eli malliin pohjautuvan ensimmäisen arvauksen ja todellisten havaintojen erotusta. Data-assimilaatioprosessin tuottama analyysi saadaan siis lisäämällä taustakenttään innovaatiot annetuilla painoilla  $W$ . Painot perustuvat mallin sekä havaintojen ennalta arvioituihin virhekovariansseihin. Erilaiset data-assimilaatiomenetelmät pohjautuvat kaikki pääpiirteissään yhtälöön 2.1, mutta ne eroavat toisistaan tulkintatavoissa; usein eroavaisuudet liittyvät painojen  $W$  määrittämiseen, kuten seuraavaksi esiteltävien empiirisen analyysin (luku 2.2) sekä tilastollisen analyysin (luku 2.3) tapauksessa. Analyysimenetelmien jakamisessa empiirisiin ja tilastollisiin menetelmiin sekä niiden esittelyssä seurataan pitkälti Eugenia Kalnayn teoksen *Atmospheric modeling, data assimilation and predictability* (2003) esittelytapaa.

## 2.2 Empiirinen analyysi

Empiriisellä analyysimenetelmällä tarkoitetaan sellaista data-assimilaatiomenetelmää, jossa havaintojen painotus  $W$  on määritelty empiirisesti tilastollisten menetelmien si-  
jaan (Kalnay, 2003). Bergthörsson ja Döös (1955) esittelivät ensimmäisinä myöhemmin  
*perättäisenä korjausmetodina* tunnetuksi tulleen empiirisen analyysimenetelmän, jossa  
havaintojen painotus yksittäisessä hilapisteessä perustuu havaintolähteiden etäisyyteen  
kyseisestä pisteestä. Perättäisen korjausmetodin lisäksi empiiristen analyysimenetel-  
mien joukkoon luetaan myös esimerkiksi ns. ”*tuuppausmetodi*”, jossa tarkoituksena on  
muokata malliyhtälöiden ratkaisu vastaamaan havaintoja empiirisen kokemuksen poh-  
jalta (Kalnay, 2003).

### 2.2.1 Perättäinen korjausmetodi

Perättäinen korjausmetodi on empiirinen analyysimenetelmä, jossa ilmakehän tila kor-  
jataan kaikissa hilapisteissä vastaamaan paremmin havaintotietoa useiden toistojen eli  
iteraatioiden kautta. Tätä perättäisten korjausten sarjaa voidaan kuvata seuraavalla  
yhtälöllä (Kalnay, 2003):

$$(2.2) \quad f_i^{n+1} = f_i^n + \frac{\sum_{k=1}^{K_i} w_{ik}^n (f_k^o - f_k^n)}{\sum_{k=1}^{K_i} w_{ik}^n + \epsilon^2}$$

jossa  $f_i^n$  on ilmakehän tila hilapisteessä  $i$  ja iteraatiossa  $n$ . Tätä korjataan vastaa-  
maan paremmin käytettävissä olevia havaintoja ( $f_k^o$ ) hyödyntämällä korjaustekijää  
 $\frac{\sum_{k=1}^{K_i} w_{ik}^n (f_k^o - f_k^n)}{\sum_{k=1}^{K_i} w_{ik}^n + \epsilon^2}$  (vrt. yhtälössä 2.1 tekijä  $W[y^o - H(x^b)]$ ), jossa  $w_{ik}^n$  on yksittäisen ha-  
vainnon  $k$  painotusfunktio hilapisteessä  $i$ ,  $\epsilon^2$  on arvio havaintolähteiden ja taustakentän  
virhevarianssien suhteesta,  $f_k^o$  on tehty havainto ilmakehän tilasta ja  $f_k^n$  arvaus ilma-  
kehän tilasta havaintopisteessä  $k$ . Tässä tapauksessa  $(f_k^o - f_k^n)$  vastaa luvussa 2.1 esi-  
teltyä innovaatiota. Perättäisen korjausmetodin ensimmäisessä iteraatiossa määrätään  
siis arvaus ilmakehän tilasta vastaamaan taustakenttää niin että

$$(2.3) \quad f_i^0 = f_i^b$$

jonka jälkeen ilmakehän tila jokaisessa hilapisteessä määrätään yhtälön 2.2 mukaisesti  
iteroimalla, kunnes haluttu tarkkuus on saavutettu. Koska kyseessä on empiirinen me-  
todi, painotusfunktio  $w_{ik}^n$  on voitu määrittää useilla tavoilla, joista tunnetuimmat ovat  
Barnesin kaava (Barnes, 1964) sekä Cressmanin kaava (Cressman, 1959). Molemmissa  
kaavoissa painotusfunktio määritetään havainnon ja hilapisteen etäisyyden avulla niin  
että:

$$(2.4) \quad w_{ik}^n = \frac{R_n^2 - r_{ik}^2}{R_n^2 + r_{ik}^2} \text{ (Cressman); } w_{ik}^n = e^{\frac{-r_{ik}^2}{2R_n^2}} \text{ (Barnes)}$$

jossa  $r_{ik}$  on hila- sekä havaintopisteen välinen etäisyys ja  $R_m$  on käyttäjän määrittämä vaikutussäde. Esitelty Cressmanin kaava pätee, kun etäisyyden  $r_{ik}$  neliö on yhtäsuuri tai pienempi kuin vaikutussäteen neliö; muussa tapauksessa  $w_{ik}^n = 0$  (Cressman, 1959). Sekä Cressmanin että Barnesin versiossa perättäisestä korjausmetodista virhetekijä määritetään yleensä niin että  $\epsilon^2 = 0$ , minkä vuoksi analyysi usein konvergoi kohti havaintoja, jos vaikutussäde on pieni (Kalnay, 2003). Koska painotusfunktion pohjaaminen pelkästään havaintolähteiden etäisyyteen hilapisteestä ei ota huomioon esimerkiksi juurikin havaintolähteiden erisuuruisia virheitä, on perättäisen korjausmetodin lisäksi kehitetty muita, tarkempia assimilaatiomenetelmiä, joista seuraavaksi esitellään tilastollisia metodeja hyödyntävät optimaalinen interpolointi ja Kalman-suodatus.

## 2.3 Tilastollinen analyysi

Painotusfunktio  $W$  voidaan määrittää empiiriseen pohjatietoon liittyvien menetelmien lisäksi tilastollisen tiedon avulla. Eräs menetelmä, johon useat tilastolliset analyysimenetelmät pohjautuvat, on *pienimmän neliösumman menetelmä*, jonka mukaan kahden tilamuuttujahavainnon  $x_1 = x_t + \epsilon_1$  ja  $x_2 = x_t + \epsilon_2$  (missä  $x_t$  on todellinen tila) avulla lasketun analyysin tarkkuudeksi saadaan (esimerkiksi Talagrand, 1997):

$$(2.5) \quad \frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

Tähän tulokseen päädytään, kun todellista tilaa arvioidaan kahden tuntemamme havainnon  $x_1$  ja  $x_2$  lineaarikombinaationa:

$$(2.6) \quad x_a = a_1 x_1 + a_2 x_2$$

jolloin analyysi  $x_a$  on paras arvio todellisesta tilasta  $x_t$ , jos minimoidaan analyysin keskineliövirhe:

$$(2.7) \quad \sigma_a^2 = E[(x_a - x_t)^2] = E[(a_1(x_1 - x_t) + a_2(x_2 - x_t))^2]$$

jossa  $E[\cdot]$  siis kuvaa odotusarvoa eli toistuvien mittausten keskiarvoa. Yhtälöstä 2.6 voidaan johtaa kertoimille  $a_1$  ja  $a_2$  yhteys  $a_1 + a_2 = 1$ , kun oletetaan analyysi harhattomaksi. Keskineliövirheen minimointi voidaan tehdä tällöin kertoimien  $a_1$  ja  $a_2$  suhteen sijoittamalla  $a_2 = 1 - a_1$ , jolloin saadaan:

$$(2.8) \quad a_1 = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2}; \quad a_2 = \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

Sijoittamalla kertoimien kaavat (2.8) yhtälöön (2.7) saadaan lopputulemaksi (2.5). Optimaalisessa tilanteessa analyysin tarkkuus on siis havaintojen tarkkuuksien summa. Vaihtoehtoinen tilastollinen tapa johtaa optimaalinen analyysi on ns. *suurimman uskottavuuden menetelmällä* (Kalnay, 2003). Tässä työssä data-assimilaatiomenetelmät ovat kuitenkin perusteltu pienimmän neliösumman avulla, joten suurimman uskottavuuden menetelmä jätetään sen nojalla tarkastelematta.

Jos palataan takaisin data-assimilaatiota kuvaavaan perusyhtälöön (2.1), havaintojen tuomaa lisäarvoa optimaalisesti painottava tekijä  $W$  voidaan kirjoittaa nyt uudelleen seuraavassa muodossa pienimmän neliösumman menetelmän mukaisesti (Kalnay, 2003):

$$(2.9) \quad W = \sigma_b^2(\sigma_b^2 + \sigma_o^2)^{-1}$$

kun tarkastellaan skalaarisuuretta yksittäisessä pisteessä (eli yhtälö 2.1 esitettynä esimerkiksi muodossa  $x_a = x_b + W[x_o - x_b]$ ). Tällöin analyysin tarkkuus ja sitä kautta virhevarianssi on:

$$(2.10) \quad (\sigma_a^2)^{-1} = [(1 - W)\sigma_b^2]^{-1} \leftrightarrow \sigma_a^2 = (1 - W)\sigma_b^2$$

Nämä tulokset pätevät siis yksinkertaiselle skalaarisuurelle, mutta niitä voidaan soveltaa myös todellisille systeemeille. Optimaalinen interpolointi (luku 2.3.1) ja Kalman-suodatus (luku 2.3.2) pohjautuvat molemmat näiden yhtälöiden käyttöön todellisissa systeemeissä.

### 2.3.1 Optimaalinen interpolointi

Siirryttäessä yksittäisestä pisteestä todelliseen systeemiin, sijaitsevat mallinnettavat suureet ja samoin analyysi kaksi- tai kolmiulotteisessa kentässä ( $\mathbf{x}_b, \mathbf{x}_a$ ), jossa ne ovat mahdollista esittää vektorina, jonka pituus on hilapisteiden ja suureiden määrän tulo (pituus =  $n$ ). Sama pätee tällöin todelliselle ilmakehän tilalle ( $\mathbf{x}_t$ ). Havaintokenttä ( $\mathbf{y}_o$ ) taas voidaan esittää vektorina, jolla on oma pituutensa  $p$ , sillä havaintokenttä ei ole mallikentän tapaan yhtenäisessä hilassa. Eliassen (1954) sekä Gandin (1965) kehittivät ensimmäisinä luvussa 2.3 esitetyjä yhtälöitä 2.9-2.10 vastaavat kaavat todelliselle monimuuttujasysteemille. Tämä yhtälöjoukko muodostaa optimaalisen interpoloinnin (OI) teoreettisen perustan.

Optimaalisen interpoloinnin tapauksessa yhtälö 2.1 voidaan esittää muodossa:

$$(2.11) \quad \mathbf{x}_a = \mathbf{x}_b + \mathbf{W}[\mathbf{y}_o - H(\mathbf{x}_b)] = \mathbf{x}_b + \mathbf{W}\mathbf{d}$$

jossa  $\mathbf{d}$  on innovaatio ja  $\mathbf{W}$  on nyt siis matriisi, jonka koko on  $n \times p$  (kutsutaan myös *vahvistusmatriisiksi*, engl. gain matrix).  $H$  taas on epälineaarinen havainto-operaattori.

Optimaalisen interpoloinnin tapauksessa yhtälöitä 2.9 ja 2.10 vastaavat muodot ovat (Kalnay, 2003):

$$(2.12) \quad \mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$$

$$(2.13) \quad \mathbf{P}_a = (\mathbf{I}_n - \mathbf{W}\mathbf{H})\mathbf{B}$$

$\mathbf{R}$  on havaintojen virhekovarianssimatriisi (koko  $p \times p$ ) ja  $\mathbf{B}$  taas taustakentän vastaava (koko  $n \times n$ ). Virhekovarianssimatriisit operoidaan *tangentti-lineaarissa* havainto-operaattorilla  $\mathbf{H}$  tai sen transpoosilla  $\mathbf{H}^T$  tarvittaviin avaruuksiin (havaintoavaruudesta malliavaruuteen transpoosin tapauksessa ja päinvastoin tavalliselle operaattorille). Vastaten yhtälöä 2.9, yhtälön 2.12 mukaan optimaalinen innovaation painotus saadaan kertomalla taustakentän virhekovarianssi (havaintoavaruudessa) systeemin kokonaisvirhekovarianssin käänteisarvolla. Jos taustakenttä on paljon havaintokenttää virheellisempi,  $\mathbf{W}$  vaikuttaa suuresti lopulliseen analyysiin, kun taas päinvastaisessa tapauksessa havaintojen lisäarvo jää pieneksi. Yhtälössä 2.13 taas sanotaan, että analyysin virhe ( $\mathbf{P}_a$ ) on taustakentän virhekovarianssi kerrottuna yksikkömatriisin (malliavaruudessa, ala-indeksi  $n$ ) ja vahvistusmatriisin erotuksella. Kokonaisuudessaan optimaalisen interpoloinnin avulla hyvälaatuiset havainnot painavat analyysissä huonoja enemmän. Yhtälössä 2.13 kuitenkin oletetaan, että muut virhelähteet ovat tarkasti tunnettuja, mikä ei täysin pidä paikkaansa (Kalnay, 2003). Näin analyysi on todellisuudessa OI-yhtälöihin verrattuna aina edes hiukan ”epäoptimaalinen”.

### 2.3.2 Kalman-suodatin

Jos luvussa 2.3.1 esiteltyä optimaalista interpolointia käyttäisi säänennustusmallissa eli ilmakehän vallitsevan tilan lisäksi myös tulevaisuuden arvioinnissa, ilmenisi OI:n käytössä todennäköisesti puutteita. Optimaalisessa interpoloinnissa ennustekentän virhekovarianssimatriisi  $\mathbf{B}$  on ennalta määrätty ja vakio, mutta todellisuudessa ennusteesiin liittyvät virheet vaihtelevat suuresti päivästä toiseen (Kalnay, 2003). Tämä otetaan huomioon Kalman-suodattimessa (KF, eng. *Kalman Filter*; Kalman 1960), jossa assimilaatio suoritetaan kahdessa vaiheessa: ensin mallin ja edeltävän analyysin avulla tehdään ennuste tulevasta ilmakehän tilasta ( $n+1$ ) ja päivitetään ennustevirhe, jonka jälkeen vahvistusmatriisin (Kalman-suodattimen tapauksessa tapana merkitä  $\mathbf{K}$ ) sekä analyysivirheen päivityksen avulla ennusteesta kehitetään uusi analyysi.

$$(2.14) \quad \mathbf{x}_n^f = \mathbf{M}_{n-1}\mathbf{x}_{n-1}^a$$

$$(2.15) \quad \mathbf{P}_n^f = \mathbf{M}_{n-1}\mathbf{P}_{n-1}^a\mathbf{M}_{n-1}^T + \mathbf{Q}_{n-1}$$

$$(2.16) \quad \mathbf{K}_n = \mathbf{P}_n^f\mathbf{H}_n^T[\mathbf{R}_n + \mathbf{H}_n\mathbf{P}_n^f\mathbf{H}_n^T]^{-1}$$

$$(2.17) \quad \mathbf{P}_n^a = [\mathbf{I} - \mathbf{K}_n\mathbf{H}_n]\mathbf{P}_n^f$$

$$(2.18) \quad \mathbf{x}_n^a = \mathbf{x}_n^f + \mathbf{K}_n[\mathbf{y}_n^o - \mathbf{H}_n\mathbf{x}_n^f]$$



Yhtälöissä 2.14 - 2.18 (Lahoz ja Schneider, 2014) alaindeksit  $n-1$  ja  $n$  kuvaavat ajanhetkeä,  $\mathbf{M}$  on lineaarinen malli, jonka avulla edellisestä analyysistä tulee uusi ennuste ja  $\mathbf{Q}$  on mallin virhematriisi. Kalman-suodattimen analyysivaiheen yhtälöt 2.16-2.18 vastaavat pitkälti OI:n perusyhtälöitä 2.11-2.13; tärkeimpänä erotuksena on päivitettävän ennustevirheen (yhtälö 2.15) hyödyntäminen. Jos jompi kumpi tai molemmat operaattoreista  $\mathbf{H}$  tai  $\mathbf{M}$  ovat epälineaarisia, puhutaan tällöin laajennetusta Kalman-suodattimesta (EKF, *Extended Kalman Filter*), jossa ennusteen virhekovarianssin laskemista varten epälineaarinen malli  $M$  linearisoidaan tangentti-lineaarisen ( $\mathbf{L}$ , Lorenz 1965) mallin avulla. Tangentti-lineaarisen matriisin koko vastaa kuitenkin käytettävän mallin vapausasteiden määrää (tavallisesti yli  $10^6$ ), mikä tekee EKF:n käytöstä kallista (Kalnay, 2003). Seuraavaksi esiteltävät parvimenetelmät (luku 2.4) tarjoavat ratkaisun ongelman kiertämiseen epälineaarisessa järjestelmässä.

## 2.4 Parvimenetelmät

Evensen (1994) esitteli tutkimuksessaan Monte Carlo -pohjaisen menetelmän Kalman-suodattimen käytölle, jossa ennustevirhe voitaisiin raskaan tangentti-lineaarisen mallin sijaan laskea hyödyntämällä ennusteparvea. Ennusteparvi  $\mathbf{x}^{b(i)}$  saadaan lisäämällä todellisiin havaintoihin satunnaisia perturbaatioita, jolloin ennusteparvi voidaan laskea kooltaan  $k$  olevasta parvesta analyyseja:  $\{\mathbf{x}^{a(i)} : i = 1, 2, \dots, k\}$  kehittämällä jokaista parven jäsentä mallin avulla. Saaduista ennusteparven jäsenistä voidaan laskea näin paras arvio ennustelle sekä sen virheelle tilastollisen keskiarvon sekä varianssin avulla:

$$(2.19) \quad \overline{\mathbf{x}^b}(t) = k^{-1} \sum_{i=1}^k \mathbf{x}^{b(i)}(t)$$

$$(2.20) \quad \mathbf{P}^b(t) = \frac{1}{k-1} \sum_{i=1}^k (\mathbf{x}^{b(i)}(t) - \overline{\mathbf{x}^b}(t))(\mathbf{x}^{b(i)}(t) - \overline{\mathbf{x}^b}(t))^T = (k-1)^{-1} \mathbf{X}^b(\mathbf{X}^b)^T$$

jossa  $\mathbf{X}^b$  on siis ennusteparven perturbaatiot  $(\mathbf{x}_i^b - \overline{\mathbf{x}^b})$  matriisimuodossa kirjoitettuna. Tällöin Kalmanin vahvistusmatriisi ja analyysivirhe voidaan laskea yhtälöiden 2.16-2.17 mukaisesti, ja uusi analyysi taas on muotoa:

$$(2.21) \quad \mathbf{x}^{a(i)} = \mathbf{x}^{b(i)}(t_a) + \mathbf{K}(\mathbf{y}^{o(i)} - H\mathbf{x}^{b(i)}(t_a))$$

jossa todellisiin havaintoihin  $\mathbf{y}^o$  on lisätty perturbaatioita. Tämänkaltaiseen ajattelu-tapaan perustui alunperin siis Evensen (1994) ja esimerkiksi Houtekamer ja Mitchell (1998) kehittämä parven Kalman-suodatin (EnKF, *Ensemble Kalman Filter*, notaatiossa seurattu pääosin Szunyogh et al. (2007) tapaa).

Parven Kalman-suodatinta on optimoitu viimeisten vuosikymmenten aikana siirtymällä kehittämään ns. *neliöjuuri-algoritmia* hyödyntäviä menetelmiä. Niissä perturboitujen

havaintojen käytöstä on luovuttu, sillä perturboitujen havaintojen käytön on todettu esimerkiksi aiheuttavan otantavirheen kasvua (Evensen, 2004). Tällöin havaintojen avulla päivitetään ainoastaan analyysiparven keskiarvoa eli:

$$(2.22) \quad \overline{\mathbf{x}}^{\mathbf{a}} = \overline{\mathbf{x}}^{\mathbf{b}} + \mathbf{K}(\mathbf{y}^{\mathbf{o}} - H\overline{\mathbf{x}}^{\mathbf{b}})$$

Aikanotaatiosta luovutaan tässä kohtaa selkeyden vuoksi. Nyt oletetaan, että analyysiparven perturbaatiot voidaan laskea ennusteparven perturbaatioista hyödyntämällä ns. *muunnosmatriisia*  $\mathbf{T}$ , jonka avulla voidaan taas laskea analyysiparvi:

$$(2.23) \quad \mathbf{X}^{\mathbf{a}} = \mathbf{X}^{\mathbf{b}}\mathbf{T} \Rightarrow \mathbf{x}^{\mathbf{a}(\mathbf{i})} = \overline{\mathbf{x}}^{\mathbf{a}} + \mathbf{X}^{\mathbf{a}}$$

Muunnosmatriisi  $\mathbf{T}$  voidaan määrittää useilla eri tavoilla, kunhan se toteuttaa ehdon, että  $\mathbf{P}^{\mathbf{a}} = (k-1)\mathbf{X}^{\mathbf{a}}(\mathbf{X}^{\mathbf{a}})^T \approx (\mathbf{I} - \mathbf{KH})\mathbf{P}^{\mathbf{b}}$ . Esimerkiksi paikallisessa Kalman-muunnossuodattimessa (LETKF, *Local Ensemble Transform Kalman Filter*; Hunt et al. (2007)) hyödynnetään muunnosmatriisina  $\mathbf{T}$  eräänlaista painotusmatriisia  $\mathbf{W}^{\mathbf{a}} = [(k-1)\tilde{\mathbf{P}}^{\mathbf{a}}]^{1/2}$ , jossa  $\tilde{\mathbf{P}}^{\mathbf{a}}$  on analyysin paikallinen virhekovarianssi *parviavaruudessa*. Tässä parviavaruudessa analyysin keskiarvo ja virhekovarianssi voidaan esittää niin että:

$$(2.24) \quad \overline{\mathbf{w}}^{\mathbf{a}} = \tilde{\mathbf{P}}^{\mathbf{a}}(\mathbf{Y}^{\mathbf{b}})^T \mathbf{R}^{-1}(\mathbf{y}^{\mathbf{o}} - \overline{\mathbf{y}}^{\mathbf{b}})$$

$$(2.25) \quad \tilde{\mathbf{P}}^{\mathbf{a}} = [(k-1)\mathbf{I} + (\mathbf{Y}^{\mathbf{b}})^T \mathbf{R}^{-1} \mathbf{Y}^{\mathbf{b}}]^{-1}$$

missä  $\mathbf{Y}^{\mathbf{b}} = H\mathbf{X}^{\mathbf{b}}$  ja  $\overline{\mathbf{y}}^{\mathbf{b}} = H\overline{\mathbf{x}}^{\mathbf{b}}$ . Tällöin malliavaruudessa

$$(2.26) \quad \overline{\mathbf{x}}^{\mathbf{a}} = \overline{\mathbf{x}}^{\mathbf{b}} + \mathbf{X}^{\mathbf{b}}\overline{\mathbf{w}}^{\mathbf{a}}$$

$$(2.27) \quad \mathbf{P}^{\mathbf{a}} = \mathbf{X}^{\mathbf{b}}\tilde{\mathbf{P}}^{\mathbf{a}}(\mathbf{X}^{\mathbf{b}})^T$$

eli  $\mathbf{w}^{\mathbf{a}(\mathbf{i})}$  on siis parvi painotusvektoreita, jotka kuvaavat millä lineaarikombinaatioilla ennusteparvesta saadaan mahdollisimman toimiva paikallinen analyysi. LETKF:n tapauksessa paikallisuudella tarkoitetaan sitä, että jokaisessa analyysiin kuuluvassa hilapisteessä assimiloidaan samanaikaisesti vain määrätyn etäisyyden sisältä löytyvät havainnot.

Myös optimaalinen interpolointi (luku 2.3.1) voidaan laajentaa parven optimaaliseksi interpoloinniksi (EnOI, *Ensemble Optimal Interpolation*, esim. Evensen (2003)), jolloin taustakentän virhekovarianssi sekä vahvistusmatriisi voidaan esittää seuraavissa muodoissa (seuraten Oke et al. (2010) notaatiota):

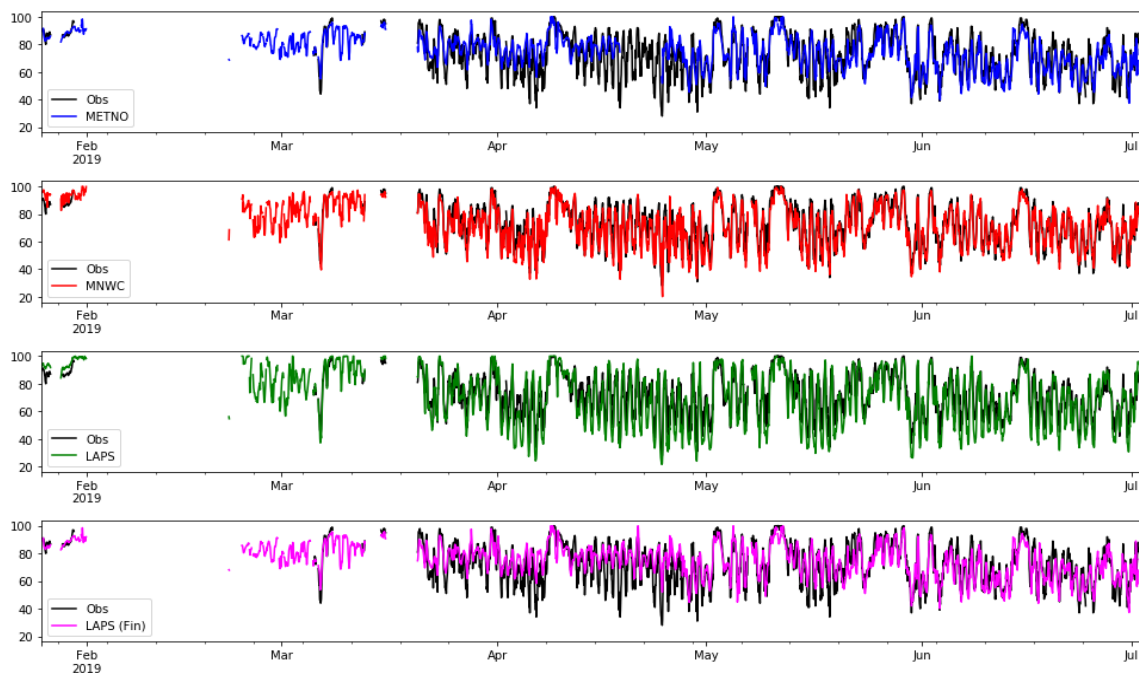
$$(2.28) \quad \mathbf{P}^{\mathbf{b}} = \rho \circ \frac{\mathbf{A}'\mathbf{A}'^T}{k-1}$$

$$(2.29) \quad \mathbf{W} = \rho \circ \mathbf{A}'(\mathbf{H}\mathbf{A}')^T (\rho \circ \mathbf{H}\mathbf{A}'(\mathbf{H}\mathbf{A}')^T + (k-1)\mathbf{R})^{-1}$$

jossa  $\mathbf{A}'$  on siis ennusteparvi  $\mathbf{A}' = \alpha[\mathbf{x}^{b(1)}, \mathbf{x}^{b(2)}, \dots, \mathbf{x}^{b(k)}]$ ,  $\alpha$  on skalaari ja  $\rho$  on korrelaatiofunktio, jonka avulla taustakentän virhekovarianssi sekä vahvistusmatriisi voidaan esittää paikallisessa muodossa. Parven käytön rajaamisella paikalliseksi on se etu erityisesti pienten parvikokojen tapauksessa, että silloin otantavirhe ei pääse kasvaamaan liian suureksi (Oke et al., 2010); toisaalta paikallisten menetelmien ongelma on se, että ne voivat tehdä analyyseistä alueellisesti epäjohdonmukaisia (Mitchell et al., 2002).

### 3. Aineisto

Tutkielmassa käytetty aineisto on peräisin karkeasti kahdesta erilaisesta lähteestä; havainnoista sekä kolmesta eri analyysi- ja nowcast-mallista. Sekä SYNOP-sääasemia (197 kpl) että kansalaishavaintoja (Netatmo-asemat, 434 kpl; lisää luvussa 3.1) Suomen alueelta käytettiin data-analyysin verrokkiaineistona. Tuntikohtaista havainto- ja mallidataa kerättiin data-analyysia varten yhteensä seitsemän eri kuukauden ajalta (25.01.2019 13 UTC - 02.07.2019 07 UTC), mutta aineiston keruussa tietokantaan oli tuntemattomasta syystä suurempi katkos ajanjakson alkupäässä (kuva 3.1). Jokainen analyysi- sekä nowcast-malli esitellään tässä luvussa erikseen; LAPS luvussa 3.2.1, norjalainen analyysi luvussa 3.2.2 ja MEPS-säämalliin pohjautuva nowcast-malli luvussa 3.2.3. Lisäksi luvussa 3.1 eritellään analyysi- ja nowcast-malleihin syötettävät havaintolähteet.



**Kuva 3.1:** Esimerkkiaikasarja, jossa on nähtävissä malliaineiston pituus. Kuvassa eri mallien suhteellisen kosteuden aikasarjat eräällä suomalaisella SYNOP-havaintoasemalla.

## 3.1 Havaintoverkostot

Analyysi- ja nowcast-mallit pystyvät hyödyntämään monipuolisesti erilaisista lähteistä kerättyä havaintotietoa, sillä esimerkiksi satelliittidatasta saatavaa kirkkauslämpötilaa voidaan käyttää pilvianalyysien teossa (Albers et al., 1996). Taulukossa 3.1 on lueteltu tässä tutkielmassa vertailtavissa analyysi- ja nowcast-malleissa käytettyjä havaintolähteitä. Tunneittain päivittyvänä nowcast-mallina MEPS-nowcast ei kykene hyödyntämään havaintoja täydessä laajuudessa, sillä rajallisesta laskenta-ajasta johtuen havaintojen latauksen lopetusaika (eng. *cut-off time*) on 15 minuuttia (luku 3.2.3). Myös METNO:n lämpötila-analyysissa hyödynnetään niukemmin eri havaintolähteitä esimerkiksi siitä syystä, että kyseessä on pelkästään pintalämpötilan analyysi, jolloin lentokone- ja luotausdataa ei tarvita analyysin muodostamiseen.

**Taulukko 3.1:** Eri analyysijärjestelmiin Suomen alueella syötettävät havainnot sekä järjestelmissä käytetyt taustakentät ja resoluutiot. Muussa kuin lämpötilan tapauksessa METNO-analyysit ovat nykyhetken MEPS-ennusteita, jolloin syötettävät havainnot vastaavat MEPS-nowcastia luotaukset kuitenkin mukaan lukien.

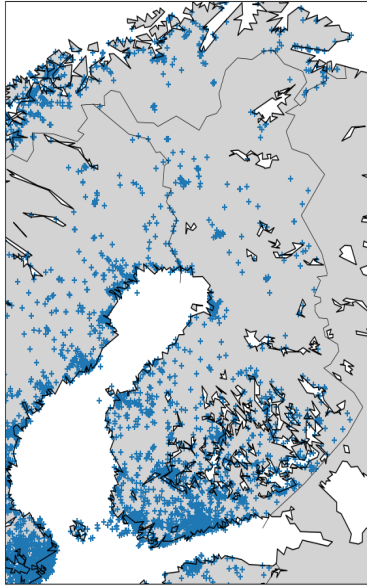
Havaintolähde	LAPS Scand/Finland	MEPS-nowcast	METNO (T)
SYNOP + METAR	Kyllä	Kyllä (osa)	Kyllä
Tiesääasemat	Kyllä	Ei	Ei
Tutkadata	Kyllä	Kyllä	Ei
Lentokonedata	Kyllä	Kyllä (osa)	Ei
Satelliittidata	Kyllä	Kyllä	Kyllä
Luotaukset	Kyllä	Eivät ehdi	Ei
Netatmo	Ei	Ei	Kyllä
Taustakenttä	MEPS/ECMWF	MEPS	MEPS-parvi
Resoluutio (vaaka)	2.5/3 km	2.5 km	1 km

Taulukon 3.1 mukaisesti ns. Netatmo-kansalaishavaintoja hyödynnetään ainoastaan norjalaisessa analyysimallissa (pelkästään lämpötila-analyysi, luku 3.2.2). Kaikille muille tutkittavista suureista ja malleista Netatmo-havainnot ovatkin riippumattomia havaintoja, ja siksi niitä käytettiin tutkielmassa myös verrokkiaineistona (luku 5.2). Netatmo viittaa siis samannimisen yrityksen valmistamiin sääasemiin (Netatmo Weather Station<sup>1</sup>), jollaisen kuka tahansa kuluttaja voi ostaa ja asentaa tahtomaansa paikkaan. Lisäksi Netatmo-asemien säätiedot lähetetään pilvipalveluun, jolloin havaintoja voi vapaasti tarkastella internetissä<sup>2</sup>. Koska Netatmo-asemat ovat kuluttajien ylläpitämiä, niiden keräämän datan laatu on hyvin vaihtelevaa ja jatkokäyttäjät joutuvat itse huo-

<sup>1</sup>url: <https://www.netatmo.com/en-us/weather>; Sivustolla käyty 23.9.2019

<sup>2</sup>url: <https://weathermap.netatmo.com>; Sivustolla käyty 23.9.2019

lehtimaan laadun varmistuksesta (luku 3.2.2 ja luku 4.1). Netatmo-sääasemat sisältävät vakiovarustuksena tutkielman kannalta kiinnostavista suureista sensorit lämpötilalle, suhteelliselle kosteudelle ja paineelle (sensorien tarkkuudet ovat Chapman et al., 2017 mukaan  $\pm 0.3^\circ\text{C}$  lämpötilalle,  $\pm 3\%$  kosteudelle ja  $\pm 1$  hPa paineelle), mutta myös sademäärää ja tuulta on mahdollista mitata lisävarusteena hankittavien sensorien avulla. Tutkielmassa verrokkiaineistona käytettyjen asemien sijainnit on esitetty kuvassa 3.2.



**Kuva 3.2:** Netatmo-asemien sijainnit Suomessa ja sen lähiympäristössä (sininen risti) piirrettynä kartalle.

## 3.2 Analyysi- ja nowcastmallit

Seuraavaksi esitellään tutkielman malliaineistot eli käydään läpi kahden analyysimallin - ns. LAPS:n (luku 3.2.1) sekä ns. METNO:n (luku 3.2.2) - ja yhden nowcast-mallin (ns. MEPS-nowcast; luku 3.2.3) erityispiirteet. Kaikkien mallien kohdalla mallien resoluutiot on tarkistettu tulostiedostojen metadatatista.

### 3.2.1 Local Analysis and Prediction System (LAPS)

Vuodesta 2009 lähtien Ilmatieteen laitoksella on ollut operatiivisessa käytössä NOAA:ssa (*National Oceanic and Atmospheric Administration*) kehitetty avoimen lähdekoodin analyysijärjestelmä, joka tunnetaan lyhenteellä LAPS (Local Analysis and Prediction System; Gregow, 2018). LAPS-analyysien kehitys aloitettiin 1980- ja 90-

lukujen vaihteessa (Birkenheuer 1999, Snook et al. 1995), ja LAPS-järjestelmän avulla on mahdollista tuottaa mesoskaalan kolmiulotteisia analyyseja suurelle joukolle meteorologisia muuttujia (McGinley et al. 1991, Albers et al. 1996, Birkenheuer 1999). LAPS pystyy hyödyntämään reaaliajassa erilaisia havaintolähteitä (lista Ilmatieteen laitoksella käytettävistä havaintolähteistä on taulukossa 3.1) ja näin tuottamaan hienohilaisen alueellisen analyysin karkeahilaisemmän taustakentän päälle (Snook et al. 1995, Gregow 2018). Tällöin siitä on hyötyä niin nowcast- kuin tutkimustarkoituksissa (Hiemstra et al., 2006); esimerkiksi Ilmatieteen laitoksella LAPS-järjestelmää on hyödynnetty muun muassa Helsinki Testbed-verkostossa (Koskinen et al., 2011).

Ilmatieteen laitoksella LAPS-järjestelmää ajetaan kahdella aluemäärittelyllä niin että Skandinavian alueen LAPS:ssa hilaväli on vaakasuunnassa 2.5 kilometriä (LAPS Scand), kun taas Suomen alueen LAPS:ssa (LAPS Finland) hilaväli on 3 kilometriä. Pystysuunnassa ilmakehä on kuvattu 44 tasolla, joiden resoluutio heikkenee mitä korkeammalle nouseaan (Gregow, 2018). Taustakenttänä Ilmatieteen laitoksen LAPS:ssa käytettiin ennen sekä Skandinavian että Suomen aluemäärittelyissä Euroopan keskipitkien ennusteiden keskuksen (ECMWF) tuottamaa viimeisintä sääennustetta, mutta nykyisin Skandinavian alueella MEPS-säämallin viimeisin ennuste on korvannut ECMWF-säämallin (enemmän MEPS-mallista luvussa 3.2.3). Havaintojen sisältämä informaatio assimiloidaan taustakenttien päälle pintasuureiden tapauksessa perättäisen korjausmetodin avulla (McGinley et al. 1991; perättäinen korjausmetodi kuvattuna luvussa 2.2.1); Albers et al. (1996) avaavat artikkelissaan laajemmin taas kolmiulotteisen analyysin taustoja.

Luvussa 2.2.1 esiteltiin kaksi tapaa ilmaista perättäisen korjausmetodin painotusfunktio, joista LAPS:ssa hyödynnetään Barnesin versiota (Barnes, 1964). Vaikka alun perin perättäiseen korjausmetodiin ei ole sisältynyt havaintovirheiden käsittelyä, LAPS-järjestelmässä virheiden vaikutus on mukana (Hiemstra et al., 2006). Lisäksi syötettävät havainnot käyvät läpi laatutarkastelun, jossa havainnot esimerkiksi verrataan klimatologiaan (Hiemstra et al., 2006) ja varmistetaan, etteivät ne eroa liikaa taustakentän arvoista (McGinley et al., 1991). Koska havaintoasemien korkeudet ja LAPS-järjestelmän maastonkuvaus eroavat toisistaan, lämpötila- sekä kastepistehavainnot korjataan keskimääräisen lämpötilavähetteen avulla; tuulelle ja painekentälle taas suoritetaan LAPS:ssa tarkasteluja, jotta niiden analyysit pysyisivät keskenään johdonmukaisina (Hiemstra et al., 2006).

### 3.2.2 MET Nordic Analysis (METNO)

MET Nordic Analysis (METNO) on Norjan ilmatieteen laitoksen kehittämä analyysijärjestelmä, joka tuottaa tällä hetkellä operatiivisesti analyysia pintalämpötilasta.

Järjestelmä tuottaa myös muiden suureiden kuten pintapaineen analyyseja, mutta ne perustuvat MEPS-mallin vallitsevan ajanhetken ennusteeseen. Tänä vuonna operatiivisen käyttöön tuodusta METNO:sta ei ole vielä olemassa virallista dokumentaatiota, joten seuraavan kuvauksen tiedot perustuvat pääosin järjestelmän kehittäjän kanssa käytyyn sähköpostiviestittelyyn (Seierstad, 2019).

METNO:n lämpötila-analyysin toiminta perustuu perinteisen yhden mallin sijaan parvimenetelmiin, sillä taustakenttänä METNO käyttää MEPS-parven (MetCoOp Ensemble Prediction System) kymmentä jäsentä. MEPS-parvi päivittyy neljä kertaa vuorokaudessa, mutta itse analyysia päivitetään kerran tunnissa uusien havaintojen avulla. Havainnot assimiloidaan taustakentän päälle parven optimaalisen interpoloinnin eli EnOI:n sekä paikallisen Kalman-muunnossuodattimen eli LETKF:n yhdistelmällä (luku 2.4). Lista METNO-analyysiin syötettävistä havainnoista on taulukossa 3.1. Taulukosta nähdään, että METNO pystyy hyödyntämään lämpötila-analyysissaan Netatmo-kansalaishavaintoja, mikä erottaa METNO-järjestelmän muista tämän tutkielman analyyseista. METNO:n lämpötila-analyysiin syötettävät Netatmo-asemien havainnot käyvät läpi muista havaintolähteistä eroavan laatutarkastelun, jossa kansalaishavaintoja verrataan joka tunnille erikseen esimerkiksi niin taustaparvikentän keskiarvoon kuin muihin lähimpiin kansalaishavaintoihin<sup>3</sup> (Nipen et al., 2015).

METNO-analyysien vaakasuuntainen resoluutio on 1 km, johon MEPS-parven tai operatiivisen MEPS-mallin karkemmat resoluutiot tuodaan. Hilan pienentäminen suoritetaan METNO:ssa hilapiste kerrallaan hyödyntäen Norjan ilmatieteen laitoksen kehittämää avoimen lähdekoodin ns. hilautettua post-prosessoria (Gridded post-processor, gridpp<sup>4</sup>), joka valitun menetelmän, kuten esimerkiksi bilineaarisen interpoloinnin, avulla suorittaa hilan pienennyksen.

### 3.2.3 MEPS-nowcast (MNWC)

Vuoden 2018 kesäkuusta lähtien Ilmatieteen laitoksella on ollut käytössä pohjoismaisen yhteistyön tuloksena kehitetyn hienohilaisen MEPS-säämallin (MetCoOp EPS eli Meteorological Cooperation Ensemble Prediction System) pohjalta luotuja nowcast-ennusteita. MEPS on lyhyen aikavälin parven säämalli, eli MEPS:n tuottamat ennusteet yltävät 66 tunnin päähän kontrolliajon tapauksessa ja 54 tunnin (päivitetty alkuperäisestä 48 tunnista) päähän parven jäsenten tapauksessa (9 kpl kontrolliajon lisäksi; Andrae, 2017). Ennen MEPS:ä lyhyen aikavälin ennustamiseen käytettiin ns. HARMONIE-säämallia. Erona oli, että HARMONIE ei tuottanut parviennusteita; MEPS on periaatteessa siis jatkokehitetty HARMONIE (verioista 40h1.1), johon

<sup>3</sup>url: <https://github.com/metno/TITAN>; Sivustolla käyty 24.9.2019

<sup>4</sup>url: <https://github.com/metno/gridpp>; Sivustolla käyty 24.9.2019



on sisällytetty parven jäsenet. HARMONIE, tai tarkemmin sanottuna HARMONIE-AROME, on taas pohjoismaisiin olosuhteisiin optimoitu versio keskieurooppalaisessa yhteistyössä kehitetystä AROME-mallista (Bengtsson et al., 2017). HARMONIE:n vaakasuuntainen hilaväli on 2,5 km, kun taas pystysuunnassa ilmakehä on kuvattu 65 tasolla ja reuna- sekä alkuehdot saadaan ECMWF-säämallista (Müller et al., 2017).

Joka kolmas tunti MEPS-säämallia päivitetään uusimmalla havaintotiedolla, mutta varsinainen pidempi ennuste tuotetaan neljä kertaa vuorokaudessa (0000,0600,1200,1800 UTC); ns. väliennusteita (0300,0900,1500,2100 UTC) käytetään seuraavissa päivityksissä taustakenttinä (Müller et al., 2017). Nämä kolmen tunnin välein tapahtuvat päivitykset taas toimivat ns. MEPS-nowcast -mallin (MNWC) taustakenttinä (Vignes ja Andrae, 2019). MNWC tuottaa varsinaisen MEPS:n tapaan kolmiulotteisia ennusteita, eli se hyödyntää samoja assimilaatiomenetelmiä kuin varsinainen MEPS-säämalli. Pinta-analyysit tuotetaan ns. CANARI-skeeman (Code d'Analyse Nécessaire à ARPEGE pour ses Rejets et son Initialization) avulla, jossa havainnot assimiloidaan hyödyntäen optimaalista interpolointia (luku 2.3), kun taas ylemmillä tasoilla assimilointiin käytetään ns. kolmiulotteista variaatioassimilointia (3DVAR) (Müller et al., 2017). Toisin kuin varsinainen MEPS-säämalli, MNWC:n 0-9h ennusteet päivittyvät joka tunti (Vignes ja Andrae, 2019); vilkkaan päivitysnopeuden vuoksi kaikkia havaintoja ei voida käyttää hyödyksi MNWC-ennusteissa (taulukko 3.1).

## 4. Menetelmät

Jotta luvussa 3 esitettyä aineistoa voitiin verrata havaintotietoon, oli käytettävä erilaisia tilastollisia lähestymistapoja data-analyysin suorittamiseen. Tässä luvussa kuvataan käytetyt tilastolliset menetelmät sekä prosessit, joiden avulla malliaineistoa sekä havaintoaineistoa voitiin verrata keskenään.

### 4.1 Tilastollinen lähestymistapa

Tämän tutkielman puitteissa verrattiin luvussa 3.2 esiteltyjen analyysijärjestelmien aineistoa käsillä olevaan havaintotietoon eli niin järjestelmään syötettäviin SYNOPSIS-pintasääs asemien havaintoihin (riippuva havaintotieto) kuin Netatmo-asemien havaintoihin (pääosin riippumaton havaintotieto, kts. luku 3.1). Vertailuun valittiin 197 SYNOPSIS-asemaa sekä 494 Netatmo-asemaa Suomen alueella, joilta löytyi havaintotietoa edes jollekin seuraavista pintaparametreista: kahden metrin lämpötila, merenpintapaine, kahden metrin suhteellinen kosteus ja tuuli kymmenessä metrissä (jatkossa tässä työssä suureiden korkeusmääritelmiä ei käytetä niistä puhuttaessa, eli kahden metrin lämpötila on pelkästään lämpötila). Käytettävissä olevan havaintoaineiston määrä vaihteli siis verifoitavan muuttujan mukaan. Tämä muuttujajoukko oli yhteinen kaikille analyysijärjestelmille, mistä syystä data-analyysi rajattiin niihin. Koska havaintotieto on tietyissä pisteissä ja analyysien mallitieto on taas määritetty erilaisiin hiloihin, oli hilamuotoiset analyysit uudelleenlaskettava näihin havaintoasemien määräämiin pisteisiin. Tässä tutkielmassa uudelleenlaskenta suoritettiin käyttämällä *lähimmän naapurin menetelmää*, eli analyysien arvoiksi havaintopisteissä määrättiin suoraan niiden hilaruutujen arvot, joissa havaintopisteet kunkin järjestelmän kohdalla sijaitsivat.

Havaintojen kerääminen tietokannasta (ns. SmartMet Serveriltä) ja analyysien uudelleenmäärittäminen lähimmän naapurin menetelmällä havaintopisteisiin oli suoritettu etukäteen tutkielmani ohjaajan FT Jussi Ylhäisin toimesta. Ylhäisi myös yhdisti havaintoaineiston ja näihin pisteisiin lasketun malliaineiston taulukoiksi, mistä eteenpäin hoidin itse varsinaisen analyysien verifiointin Python-ohjelmointikielellä (versio 3). Analyysien uudelleenlaskemisessa havaintopisteisiin käytettiin pyTroll-ohjelmiston pa-

kettia *pyresample*<sup>1</sup>, joka on kehitetty erityisesti paikkatiedon käsittelemistä ja uudelleenlaskemista varten.

Tutkimuksessa hyödynnettiin pääasiassa kahta tilastollisista apuvälinettä tulosten tulokinnassa - lineaarista regressioanalyysia tutkimaan analyysien riippuvuutta havaintotiedosta sekä keskineliövirheen neliöjuurta (RMSE, Root Mean Square Error) tilastollisena tunnuslukuna arvioimaan analyysien poikkeamia havaintotiedosta. Data-analyysissa käytettiin yksinkertaista yhden selittäjän lineaarista regressiomallia, jossa kahden muuttujan (havainto- sekä mallitiedon) välinen yhteys on kuvattavissa regressiosuoran avulla:

$$(4.1) \quad y_i = a + bx_i + \epsilon_i$$

niin että  $y_i$  on selitettävä muuttuja (tässä tutkielmassa analyysijärjestelmien suuret),  $x_i$  taas selittävä muuttuja (eli havainnot),  $a$  on y-suoran leikkauspiste,  $b$  on regressiosuoran kulmakerroin ja  $\epsilon$  mallin jäännösvirhe. Pienimmän neliösumman menetelmän avulla mallin jäännösvirhe voidaan minimoida, jolloin saadaan muuttujien väliselle riippuvuudelle mahdollisimman tarkka lineaarinen sovitus. Lisäksi  $r^2$ -luvulla eli Pearsonin korrelaatiokertoimen neliöllä tarkasteltiin, kuinka suuren osan analyysijärjestelmien parametrien vaihtelusta voi selittää havaintotiedolla. Pearsonin korrelaatiokerroin lasketaan:

$$(4.2) \quad r_{xy} = \frac{s_{xy}}{s_x s_y}$$

jossa  $s_{xy}$  on aineiston otoskovarianssi ja  $s_x$  sekä  $s_y$  tutkittavien muuttujien keskihajonnat. Analyysien poikkeamien suuruutta kuvaava RMSE taas lasketaan seuraavasti:

$$(4.3) \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2}$$

$r^2$ - sekä RMSE-luvut määritettiin sekä kokonaisille mittaussarjoille että pelkälle vuorokausivaihtelulle. Vuorokausivaihtelu tarkoitti tässä tutkimuksessa sitä, kuinka hyvin analyysit pystyivät toisintamaan havaitut *tuntikohtaiset* ääriarvot (07-18 UTC tuntiarvoista arvioidut maksimit; 19-06 UTC tuntiarvoista arvioidut minimi lämpötilalle; muilla parametreilla vuorokauden tuntiarvojen osalta arvioidut minimi ja maksimit). Lisäksi samoja tunnuslukuja laskettiin jokaiselle kuukaudelle erikseen parametreittain sekä erottaen päivä- ja yöajan arvot (liitteet A.1 sekä A.2); maantieteellisten ongelmakohtien tarkastelua varten taas määritettiin RMSE-luvut yksitellen jokaiselle havaintoasemalle. Netatmo-asemilta (luku 3.1) saadulle laatutarkastamattomalle havaintotiedolle taas suoritettiin järjestykseen tarkasteluja; jos lämpötila erosi 15 asteella tai kosteus yli 40% analyysista jätettiin kyseiset havainnot pois data-analyysista.

<sup>1</sup>url: <https://github.com/pytroll/pyresample>; Sivustolla käyty 24.9.2019

## 4.2 Kastepisteen ja merenpintapaineen laskentakaavat

Ennen kuin eri analyysieja pystyttiin vertailemaan keskenään, analyysien painelaskelmista piti ensin tehdä yhteismitallisia, sillä LAPS ja MEPS-nowcast tuottavat analyysit pintapaineesta, ja havaintotieto sekä norjalainen analyysi ilmaisevat paineen merenpintapaineena. Merenpintapaineiden laskemiseen pintapaineista käytettiin seuraavaa kaavaa (kuten Ruotsin Ilmatieteen laitoksella, lähde: World Meteorological Organization, 2012):

$$(4.4) \quad P = P_s \cdot \exp\left[\frac{H \cdot 0.034163(1 - 0.0026373\cos(2\varphi))}{T_1}\right]$$

jossa  $T_1$  käytetään kuvaamaan talvisia inversiotilanteita:

$$t < -7^\circ\text{C} \Rightarrow T_1 = 0.5 \cdot t + 275$$

$$-7^\circ\text{C} \leq t < +2^\circ\text{C} \Rightarrow T_1 = 0.535 \cdot t + 275.6$$

$$t \geq +2^\circ\text{C} \Rightarrow T_1 = 1.07 \cdot t + 274.5$$

ja  $P$  = merenpintapaine,  $P_s$  = pintapaine,  $H$  = havaintopisteen korkeus (tässä tapauksessa korkeus analyysin hilapisteessä) ja  $\varphi$  = havaintopisteen sijainti leveyspiireissä.

Merenpintapaineen laskemisen lisäksi kosteusanalyysien onnistumisen tulkintaa varten päätimme ohjaajani kanssa laskea tulokset myös kastepisteelle. Kastepistelämpötilat laskettiin hyödyntämällä ns. Magnusen kaavaa (esim. Lawrence, 2005) muistamalla että  $\frac{RH}{100} = \frac{e}{e_s}$  jolloin:

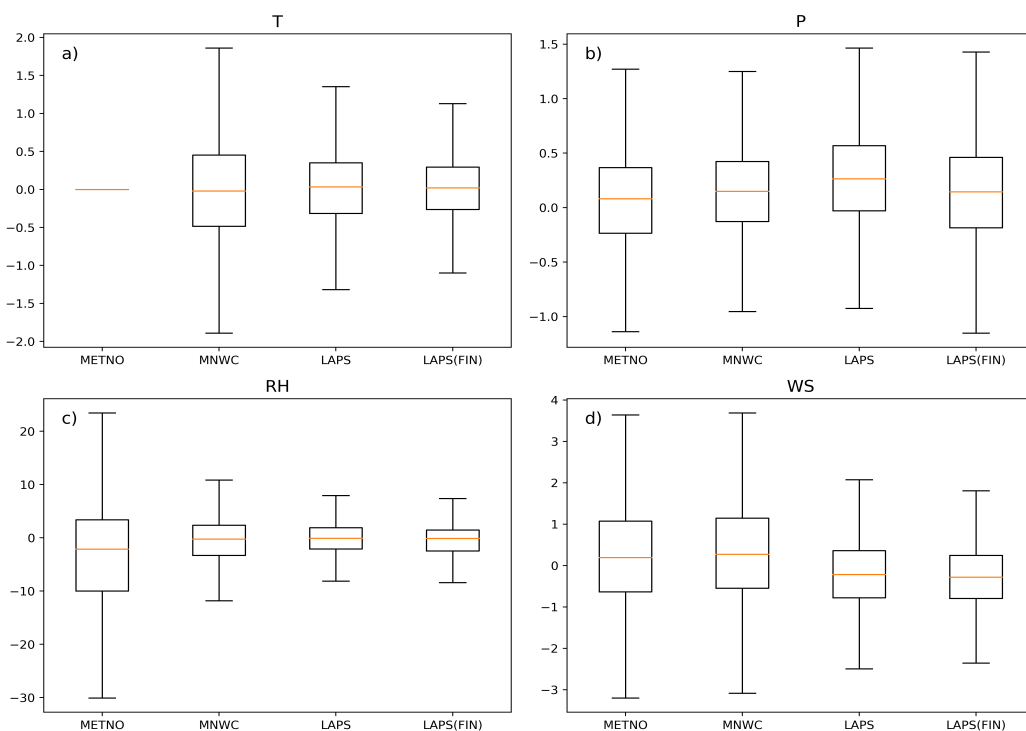
$$(4.5) \quad T_d = \frac{c \cdot \left(\ln\left(\frac{RH}{100}\right) + \frac{b \cdot T}{c+T}\right)}{b - \left(\ln\left(\frac{RH}{100}\right) + \frac{b \cdot T}{c+T}\right)}$$

jossa vakoille  $b$  ja  $c$  on käytetty arvoja  $b = 18.678$  sekä  $c = 257.14^\circ\text{C}$  (Buck, 1996), ja  $RH$  on siis suhteellinen kosteus ja  $T$  lämpötila. Sekä merenpintapaineen että kastepisteen laskemista varten tarvittiin mallien asemakohtaiset korkeudet, jotka johdettiin mallien käyttämistä korkeuskentistä aiemmin kuvatulla lähimmän hilapisteen menetelmällä.

## 5. Tulokset

Lämpötila-analyysien (luku 5.1.1), merenpintapaine-analyysien (luku 5.1.2), kosteusanalyyysien (luku 5.1.3) sekä tuulianalyysien (luku 5.1.4) verifiointitulokset riippuvaa havaintoverkostoa eli suomalaisia SYNOP-pintasääasemia vasten ovat luvun 5.1 pääsisältö. Lisäksi luvussa 5.2 käydään läpi, miten lämpötila- ja kosteusanalyyysit vertautuivat kansalaisten omistamien Netatmo-asemien havaintotietoon.

### 5.1 Pintasuureiden vertailu riippuvassa havaintoverkostossa



**Kuva 5.1:** Tilastollisten virheiden ( $\epsilon = x_{havaintu} - x_{analyysi}$ ) jakaumat esitettynä jokaiselle parametrille ja analyysijärjestelmälle erikseen. Kuvattuna on a) lämpötilan, b) paineen, c) suhteellisen kosteuden ja d) tuulen nopeuden virheet. Virheet (y-akseli) ovat esitetty ruutu- ja janakuviaina, joissa mediaani on esitetty punaisen viivan avulla, 25 ja 75-persenttiilit ruutujen ala- ja ylärajojen avulla ja maksimit sekä minimit janojen avulla. X-akselilla on kussakin kuviossa käytetty analyysijärjestelmät.

### 5.1.1 Lämpötila

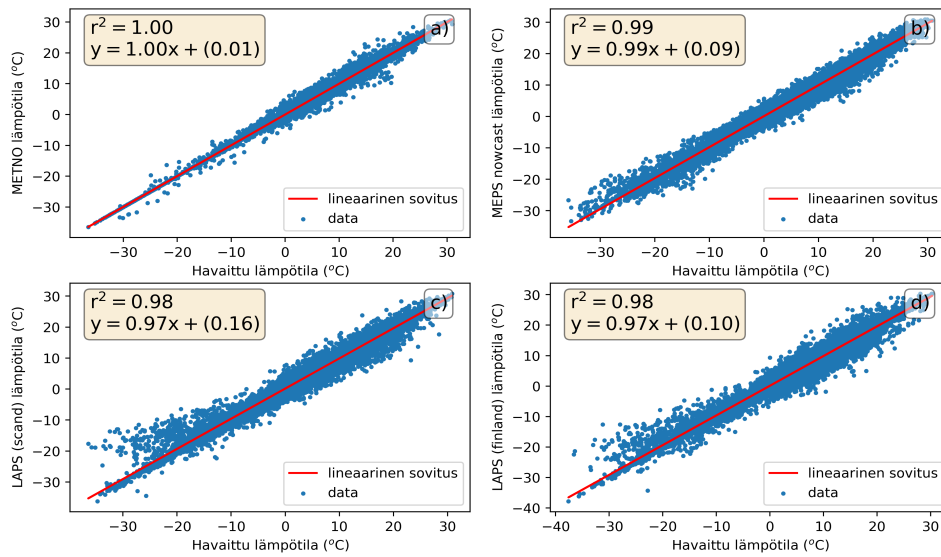
Havaittujen sekä analysoitujen lämpötilojen väliset korrelaatiot olivat tässä tutkimuksessa korkeita kaikille analyysijärjestelmille, sillä korrelaatiokertoimien neliöiden ( $r^2$ ) vaihteluväli oli 0.98 - 1.00 kahden desimaalin tarkkuudella (kuva 5.2). Taulukossa 5.1 on eriteltynä korrelaatiot ja keskineliövirheen keskijuuret kaikille analyyseille erikseen; lisäksi taulukossa on eritelty samat parametrit lämpötilan ääriarvojen tarkastelulle. Norjan ilmatieteen laitoksen analyysi METNO oli analyyseista tarkin, sillä sen korrelaatiokerroin oli pyöristettynä 1.00. Tällöin lineaarinen sovitus vastasi 1:1-suoraa (kuva 5.2a), ja RMSE oli pieni (taulukko 5.1). Muilla analyysijärjestelmillä lämpötila-analyysin korrelaatio oli tässä työssä 0.98-0.99, ja RMSE:t olivat 0.7°C tuntumassa (taulukko 5.1). Myös näiden analyysijärjestelmien kohdalla lineaariset sovitukset olivat lähellä 1:1-suhdetta, ja virheet olivat lämpötila-analyyseissa pieniä (kuva 5.1).

Sekä taulukko 5.1 että kuva 5.1a osoittavat METNO-analyysin virheiden olleen melkein mitättömiä, kun taas nowcast-mallilla virheiden vaihteluväli oli suurin. Kuvan 5.1 suhteen on kuitenkin huomioitava, että selkeyden vuoksi janakuvioissa ei ole mukana suurimpia virhepoikkeamia mediaanista (voivat olla jopa  $\pm 15$  astetta), eli todellisuudessa METNO-analyysillä oli myös nollasta poikkeavia virheitä (nähtävissä hajontakuvioista 5.2a). Päivä- ja yökohtaisen vaihtelun sekä kuukausittaisen vaihtelun tapauksessa huomattavin poikkeama oli huhtikuun lämpötila-analyysien osuvuudessa, jolloin analyysien keskivirheessä oli nähtävissä kasvua (liitteet A.1 ja A.2; taulukko A.1 ja taulukko A.6).

Analysoitujen lämpötilojen RMSE:t on esitetty jokaiselle havaintoasemalle erikseen kuvassa 5.3, johon on otettu mukaan hajontakuvioista poiketen vain toinen versio LAPS-analyysista eli Skandinavian alueen LAPS. Suurimmat virheet vaikuttavat lämpötila-analyyseissa keskittyneen osalle Lapin asemista. LAPS-analyysin tapauksessa myös osalla merellisistä asemista virheet olivat suhteellisesti suurempia.

**Taulukko 5.1:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) lämpötilan suhteen vertailukohtana 197 suomalaisen pintasääaseman havainnot.

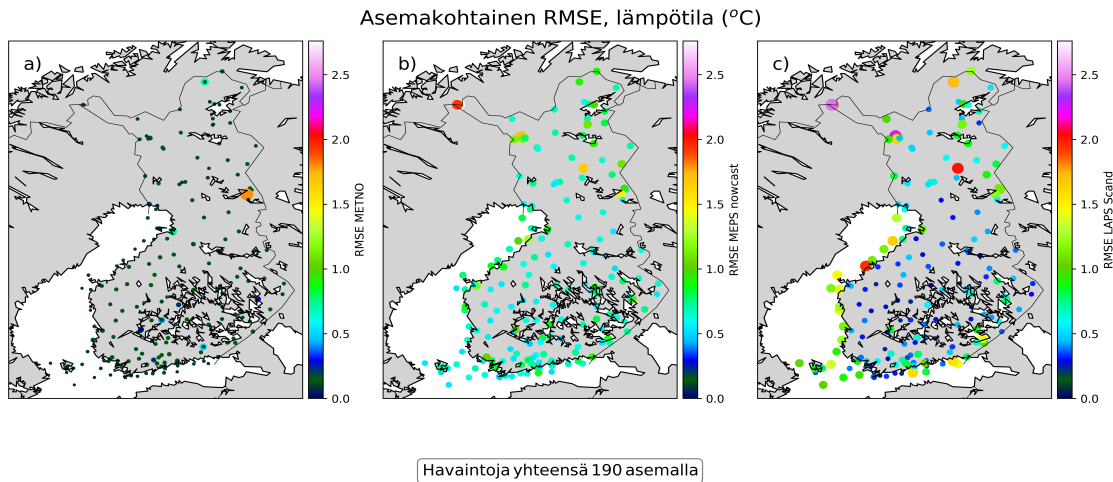
	$r^2$	RMSE (celsius)	$r^2$ (ääriarvot)	RMSE (ääriarvot)
T (METNO)	0.997	0.136	0.999	0.072
T (MNWC)	0.985	0.723	0.987	0.764
T (LAPS Scand)	0.981	0.667	0.983	0.694
T (LAPS Finland)	0.981	0.648	0.982	0.696



**Kuva 5.2:** Havaittujen lämpötilojen ja analyysijärjestelmien laskemien lämpötilojen väliset korrelaatiot esitettynä hajontakuvioiden avulla. Hajontakuviot ovat järjestyksessä a) MET Nordic Analysis, b) MEPS-nowcast, c) LAPS (Skandinavia), d) LAPS (Suomi). Kuvioissa punainen käyrä edustaa lineaarista sovitusta, jonka yhtälö on kuvattuna jokaiselle hajontakuviolle erikseen tekstilaatikossa; siniset pisteet taas ovat yksittäisiä datapisteitä.

### 5.1.2 Merenpintapaine

Korrelaatiot analysoitujen merenpintapaineiden sekä havaittujen merenpintapaineiden välillä olivat vielä suurempia kuin lämpötilan tapauksessa ( $r^2 = 1.00$  kaikille analyysijärjestelmille; kuva 5.4). Analyyseista lasketut merenpintapaineet (laskettu kuten luvussa 4.2) toisinsivat siis hyvinkin tarkasti havaitut merenpintapaineet, ja niiden suhdetta kuvaavat lineaariset sovitukset vastasivat pitkälti 1:1-suoraa. Tulokset kokonaisille mittaussarjoille ja vuorokausivaihtelulle on kuvattu tarkemmin taulukossa 5.2 ja päivä- ja yökohtaiset sekä kuukausittaiset tulokset liitteissä A.1 sekä A.2 (taulukot A.2 ja A.6).



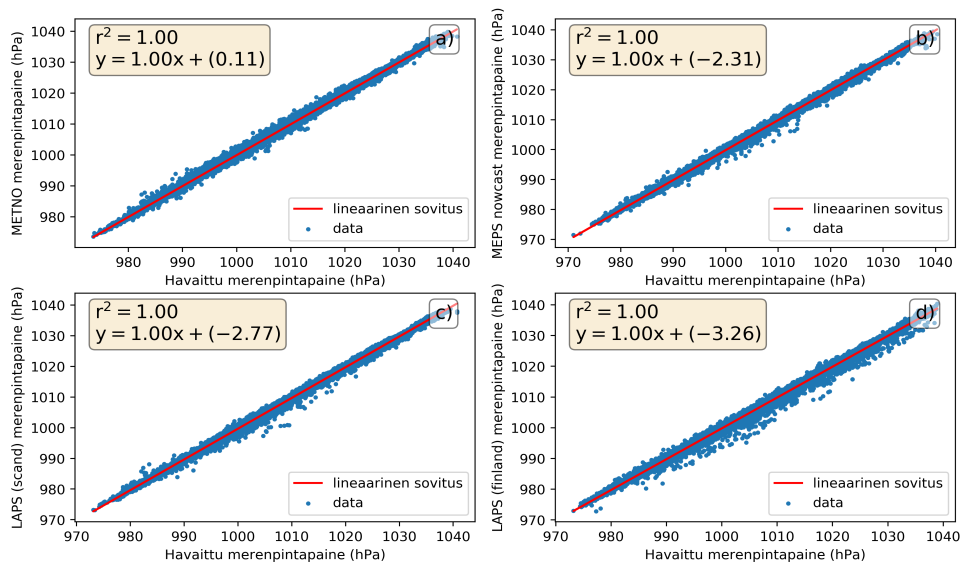
**Kuva 5.3:** Analyysijärjestelmien laskemien lämpötilojen keskineliövirheen neliöjuuret (RMSE) ilmoitettuna jokaiselle mittausjoukossa mukana olleelle lämpötilaa havaitsevalle sääasemalle erikseen. Sekä pisteiden väri että koko kuvaavat RMSE:n suuruutta mutta eri tavoin - väriskaala on kaikille malleille yhteinen kun taas pisteiden koko ilmaisee kuinka pisteen virheen suuruus suhteutuu kyseisen järjestelmän muiden pisteiden virheiden suuruuteen (koon määrää siis 0-1 välille normitetut RMSE-arvot). Kuvattuna on a) MET Nordic Analysis, b) MEPS-nowcast ja c) LAPS (Skandinavia). Lämpötilahavaintoja on kerätty yhteensä 190 asemalla 197 mukaan valikoituneesta asemasta.

**Taulukko 5.2:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) merenpintapaineen suhteen vertailukohtana 197 suomalaisen pintasääaseman havainnot.

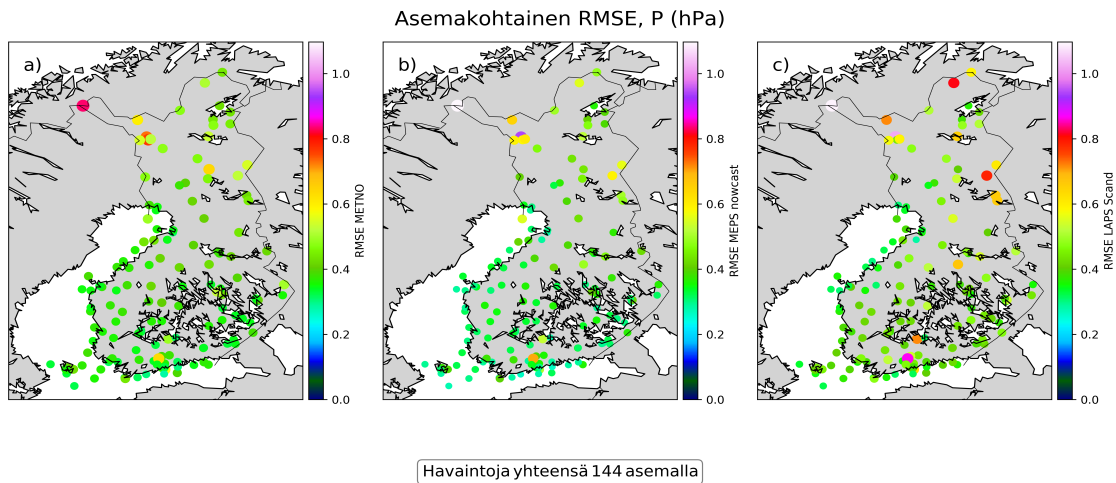
	$r^2$	RMSE (hPa)	$r^2$ (ääriarvot)	RMSE (ääriarvot)
P (METNO)	0.998	0.400	0.998	0.388
P (MNWC)	0.998	0.376	0.999	0.378
P (LAPS Scand)	0.998	0.443	0.998	0.404
P (LAPS Finland)	0.997	0.450	0.996	0.467

Kuvassa 5.5 on esitetty RMSE:t jokaiselle merenpintapainetta havainneelle asemalle erikseen samalla tavalla kuin kuvassa 5.3 lämpötilalle. Eri analyysijärjestelmien suurimmat virheet näyttävät sijoittuneen lämpötilan tapaan Pohjois-Suomeen. Toisaalta sielläkin RMSE:t olivat suurimmillaan 1,2 hPa tienoilla, eli suuria virheitä ei merenpintapaineen analysoinnissa ollut. Kaikkien analyysien tapauksessa analysoidut merenpintapaineet olivat keskimäärin hieman havaittuja paineita pienempiä (kuva 5.1b).





Kuva 5.4: Sama kuin kuva 5.2 mutta esitettynä merenpintapaineelle.



Kuva 5.5: Sama kuin kuva 5.3 mutta kuvattuna on merenpintapaineen RMSE:n maantieteelliset jakaumat. Merenpintapainehavaintoja on kerätty yhteensä 144 asemalla 197 mukana olleesta asemasta.

### 5.1.3 Suhteellinen kosteus ja kastepiste

Analyysijärjestelmien onnistumista kosteuden tulkinnessa tutkittiin ottamalla käsittelyyn sekä järjestelmien suoraan määrittelemä suhteellinen kosteus että laskemalla suhteellisen kosteuden ja lämpötilan avulla analyysien kastepistelämpötilat (kuvattu luvussa 4.2). Verrattuna lämpötila- ja paineanalyysihin ilmeni kosteuden analyysissa enemmän vaihtelua (taulukko 5.3 (RH) sekä taulukko 5.4 ( $T_d$ )) - METNO-analyysin

korrelaatiot olivat selkeästi huonommat kuin muilla järjestelmillä (RH:  $r^2 = 0.65$  vs.  $r^2 \approx 0.91$ ;  $T_d$ :  $r^2 = 0.9$  vs.  $r^2 \approx 0.98$ ), ja myös tilastollisen virheen suuruuden vaihteluväli oli selkeästi muita suurempi (kuva 5.1c).

Järjestelmistä LAPS-analyysin Skandinavian alueen versio toisti parhaiten havaitut kosteuden muutokset valituissa pisteissä, sillä niin RMSE-arvot olivat kyseisellä analyysillä pienimmät (n. 3.5% suhteelliselle kosteudelle ja  $0.6^\circ\text{C}$  kastepisteelle) kuin korrelaatio oli suurin havaittuun kosteuteen verrattuna (yli 0.9 molemmille parametreille). Taulukossa 5.3 huomioitavaa on esimerkiksi se, että RH:n tapauksessa vuorokausivaihtelu oli kuvattu tuntikohtaista vaihtelua paremmin. Myös kosteuden päivä- ja yökohtaista vaihtelua sekä kuukausittaista vaihtelua tutkittiin, ja tulokset ovat nähtävissä liitteiden A.1 ja A.2 taulukoissa A.3, A.4 ja A.6; lämpötilan tapaan (luku 5.1.1) huhtikuu osoittautui muita kuukausia ongelmallisemmaksi, ja lisäksi yöllisten kosteusanalyysien keskivirheet olivat päivää pienempiä.

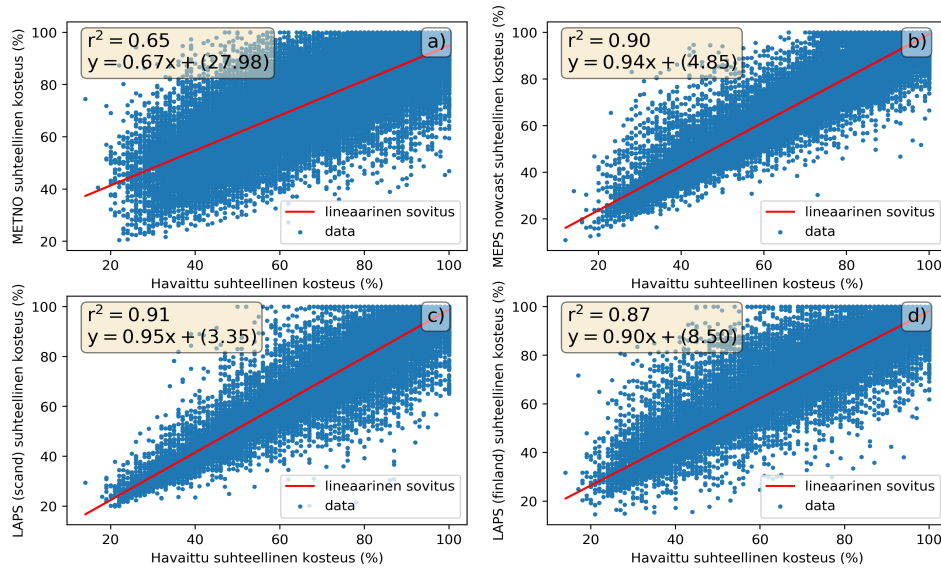
Kuten kuvasta 5.6 voidaan havaita, suhteellinen kosteus on parametri, jossa havaittujen ja analysoitujen kosteuksien välillä oli enemmän hajontaa kuin esimerkiksi merenpintapaineen yhteydessä (kuva 5.4). Kastepisteen tapauksessa taas hajontaa oli suhteellista kosteutta vähemmän (kuva 5.8). Kuvassa 5.7 on esitetty esimerkiksi kuvan 5.5 tapaan RMSE:t jokaiselle RH:ta havainneelle asemalle erikseen. Kuva 5.7 osoittaa kuvan 5.6 tapaan, että METNO-analyysillä oli eniten ongelmia suhteellisen kosteuden esittämisessä valituilla asemilla, kun taas LAPS-analyysillä virheet olivat erityisesti maa-asemilla pienimpiä.

**Taulukko 5.3:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) suhteellisen kosteuden suhteen vertailukohtana 197 suomalaisen pintasääaseman havainnot.

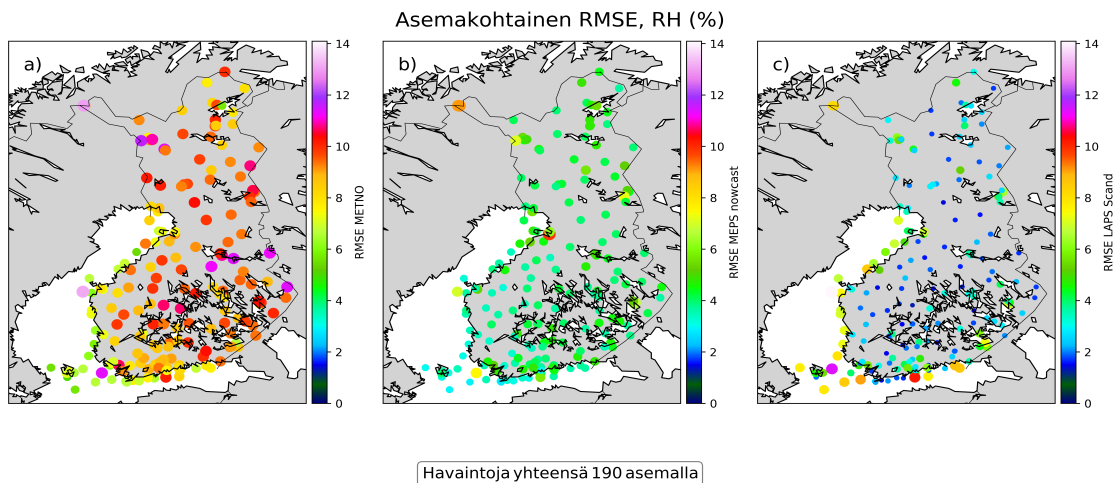
	$r^2$	RMSE (%)	$r^2$ (ääriarvot)	RMSE (ääriarvot)
RH (METNO)	0.645	8.729	0.811	7.761
RH (MNWC)	0.897	4.203	0.945	3.800
RH (LAPS Scand)	0.911	3.513	0.955	3.024
RH (LAPS Finland)	0.873	3.999	0.926	3.690

**Taulukko 5.4:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) kastepistelämpötilan suhteen vertailukohtana 197 suomalaisen pintasääaseman havainnot.

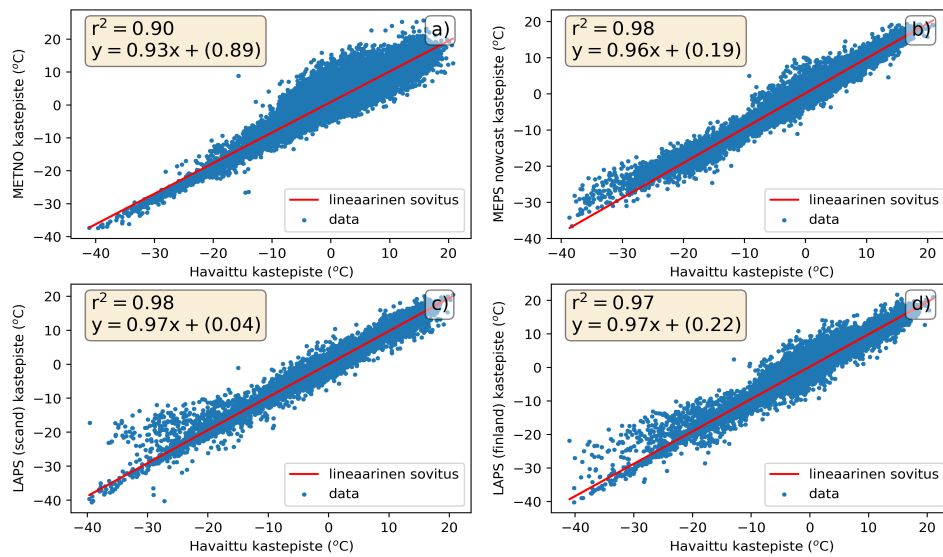
	$r^2$	RMSE (celsius)	$r^2$ (ääriarvot)	RMSE (ääriarvot)
$T_d$ (METNO)	0.900	1.824	0.924	1.795
$T_d$ (MNWC)	0.977	0.762	0.976	0.860
$T_d$ (LAPS Scand)	0.980	0.603	0.982	0.625
$T_d$ (LAPS Finland)	0.970	0.739	0.971	0.819



**Kuva 5.6:** Sama kuin kuva 5.2 mutta esitettynä suhteelliselle kosteudelle.



**Kuva 5.7:** Sama kuin kuva 5.3 mutta kuvattuna on kosteuden RMSE:n maantieteellinen jakauma. Kosteushavaintoja on kerätty yhteensä 190 asemalla 197 mukana olleesta asemasta.



**Kuva 5.8:** Sama kuin kuva 5.2 mutta esitettynä kastepistelämpötilalle.

### 5.1.4 Tuulen nopeus

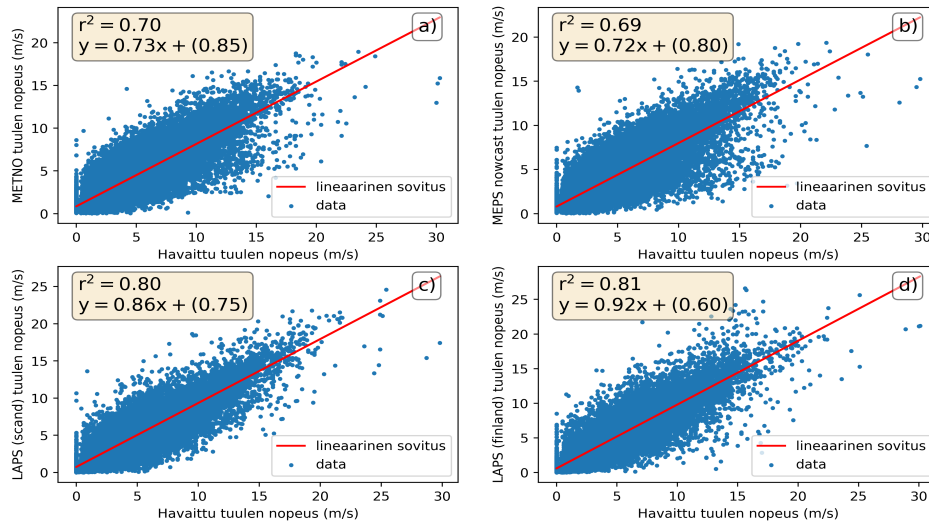
Tuulen nopeuden suhteen analysoitujen ja havaittujen nopeuksien väliset korrelaatiokertoimet osuivat pyöristettynä arvojen 0.7-0.8 välille niin, että korrelaatio oli korkein LAPS-analyysin tapauksessa (0.80-0.81 Suomen ja Skandinavian alueen LAPS:lle) ja heikoin MNWC:n tapauksessa (0.69; taulukko 5.5). Samansuuntainen tulos oli nähtävissä myös virheiden tapauksessa, sillä LAPS-analyysille RMSE oli keskimäärin 0.9 m/s, kun taas MNWC:lle RMSE oli 1.2 m/s, ja LAPS-analyysien virheiden suuruuden vaihteluvälit olivat muita pienemmät (kuva 5.1d).

Pelkkiin korrelaatiokertoimiin nojaten vaikuttaa siis siltä, että analyysijärjestelmillä oli parametreista keskimäärin eniten ongelmia havaittujen tuulen nopeuksien muutoksien toisintamisessa. Taulukosta 5.5 nähdään lisäksi, että tuulen nopeuden tapauksessa korrelaatiokertoimet olivat suurempia ja RMSE:t pienempiä tarkasteltaessa pelkästään ääriarvojen kuin koko mittausarjan toisintamista valituilla asemilla. Öisin tuulen nopeuden keskivirheiden huomattiin pienenevän (liite A.1; taulukko A.5), mutta kuukausittaisessa vaihtelussa ei havaittu selkeää trendiä (liite A.2; taulukko A.6).

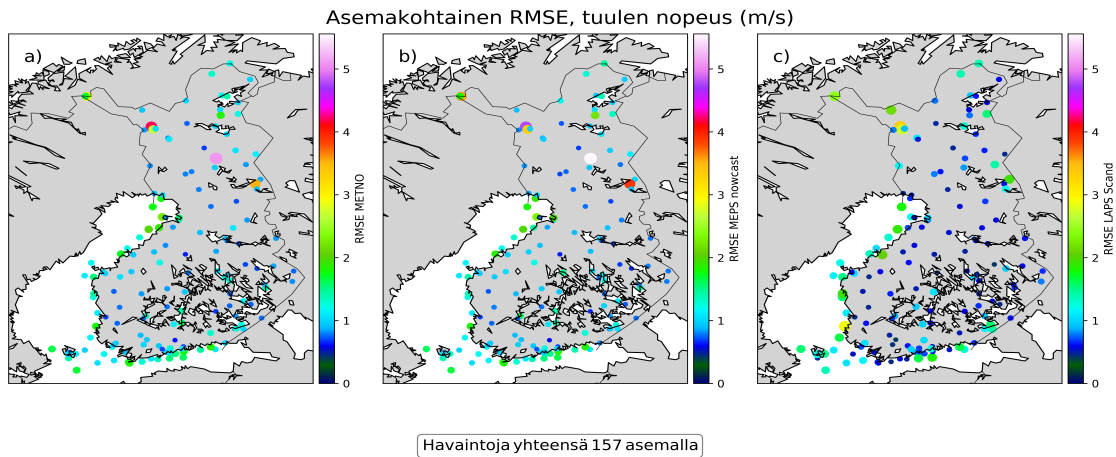
Kuvan 5.9 lineaaristen sovitusten mukaan jokaiselle tuulianalyysille oli yhteistä se, että pienillä havaituilla tuulen nopeuksilla (0-5 m/s) tuulianalyysien arvot olivat enemmän liian suuria kuin pieniä, mutta suuremmilla nopeuksilla (yli 10 m/s) tilanne kääntyi päinvastaiseksi. Maantieteellisesti suurimmat virheet tuulianalyysissä olivat osalla Lapin asemista (kuva 5.10).

**Taulukko 5.5:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) tuulen nopeuden suhteen vertailukohtana 197 suomalaisen pintasääaseman havainnot.

	$r^2$	RMSE (m/s)	$r^2$ (ääriarvot)	RMSE (ääriarvot)
WS (METNO)	0.702	1.163	0.822	1.087
WS (MNWC)	0.686	1.180	0.810	1.107
WS (LAPS Scand)	0.802	0.897	0.883	0.814
WS (LAPS Finland)	0.807	0.894	0.889	0.797



**Kuva 5.9:** Sama kuin kuva 5.2 mutta esitettynä tuulen nopeudelle.



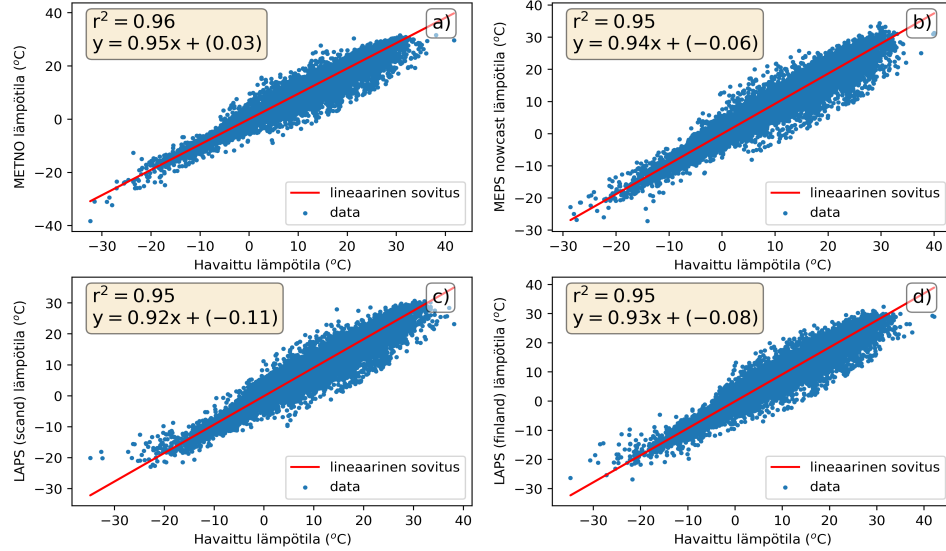
**Kuva 5.10:** Sama kuin kuva 5.3 mutta tuulen nopeudelle. Tuulihavaintoja on kerätty yhteensä 157 asemalla 197 mukana olleesta asemasta.

## 5.2 Pintasuureiden vertailu Netatmo-asemien avulla

Riippuvan havaintoverkoston eli Ilmatieteen laitoksen ylläpitämien SYNOP-havainto-asemien lisäksi pintasuureanalyysseja oli mahdollista verrata kansalaisten omistamien Netatmo-asemien (luku 3.1) havaintotietoihin. Verifointiin valikoitui Suomen alueelta 434 havaintoasemaa. Näin analyysseja pystyttiin vertaamaan riippumatonta havaintoverkostoa vasten ainoana poikkeuksena METNO-järjestelmän tuottama lämpötila-analyysi, jossa Netatmo-asetat ovat yksi käytettävistä havaintolähteistä (luku 3.1). Johtuen Netatmo-asemien havaintotiedon laadunvarmistuksen puutteesta tulosesittelyyn päätyivät suureista vain lämpötila ja kosteus. Näiden suureiden, toisin kuin tuulen nopeuden ja paineen, kohdalla oli työn laajuuden kannalta kohtuullisen helppoa suorittaa järkevyytarkasteluja Netatmo-asemien havaintosarjoille (paineen ja tuulen nopeuden hajontakuviot nähtävissä liitteissä; kuva A.5 ja kuva A.4).

Kuten riippuvan havaintoverkoston tapauksessa, myös Netatmo-asemien havaintotietoon verrattaessa lämpötila-analyysit korreloivat hyvin havaintoaineiston kanssa (kuva 5.11). Lämpötila-analyysseista METNO-järjestelmän analyysi korreloi vahviten Netatmo-asemien lämpötilahavaintojen kanssa ( $r^2 = 0.96$ ), mutta muillakin järjestelmillä korrelaatio oli korkea ( $r^2 = 0.95$ ). Sekä kokonaisten mittausarjojen että ääriarvotarkastelun RMSE:t taas olivat lämpötilan osalta 1-1.6 celsiusasteen tienoilla Netatmo-asemiin verrattaessa (taulukko 5.6). Lisäksi kuvasta 5.11 nähdään, että havaitut lämpötilat olivat keskimäärin hieman analysoituja lämpötiloja korkeampia.

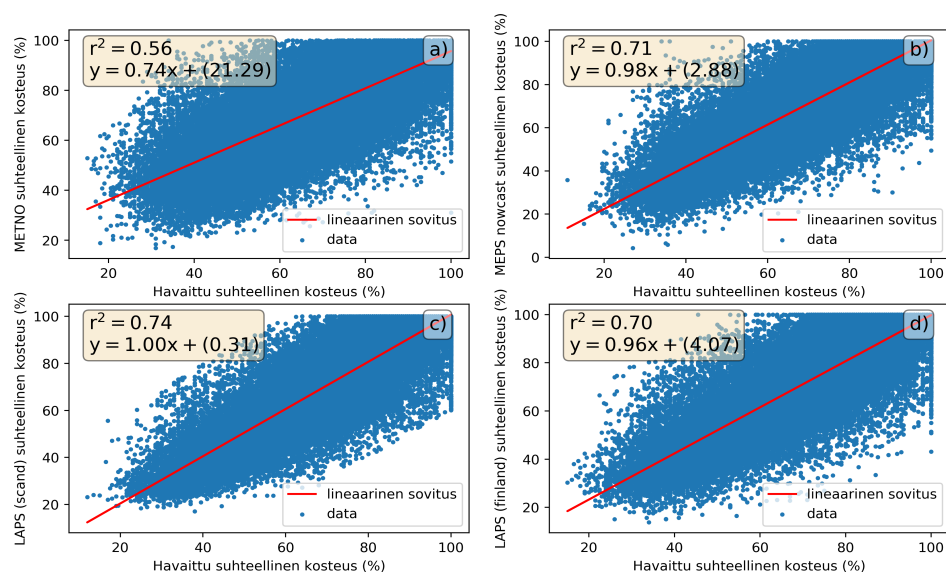
Tarkastellessa kosteusanalyysseja Netatmo-havaintoja vasten olivat tulokset melko samansuuntaisia kuin SYNOP-asemavertailuissa. METNO-analyysin suhteellinen kosteus korreloi vähiten Netatmo-asemien kosteuden kanssa ( $r^2 = 0.56$ ), kun taas LAPS-analyysin (Skandinavia) suhteellinen kosteus korreloi parhaiten havaittujen kosteuksien kanssa ( $r^2 = 0.74$ ); tämä on nähtävissä myös hajontakuvioiden muodoista (kuva 5.12). Taulukossa 5.7 on tarkemmin eritelty korrelaatiokertoimet ja RMSE-arvot kosteusanalyysseille. Liittessä A.3 on lisäksi nähtävissä Netatmo-vertailuun pohjautuvia lisätuloksia, kuten SYNOP-asemavertailussa esiteltyjä RMSE-arvojen maantieteellisiä jakaumia vastaavat kuvat Netatmo-vertailulle (kuvat A.1 ja A.2).



**Kuva 5.11:** Netatmo-asemien havaittujen lämpötilojen ja analyysijärjestelmien laskemien lämpötilojen väliset korrelaatiot esitettynä hajontakuvioiden avulla. Hajontakuviot ovat järjestyksessä a) MET Nordic Analysis, b) MEPS-nowcast, c) LAPS (Skandinavia), d) LAPS (Suomi). Kuvioissa punainen käyrä edustaa lineaarista sovitusta, jonka yhtälö on kuvattuna jokaiselle hajontakuviolle erikseen tekstilaatikossa; siniset pisteet taas ovat yksittäisiä datapisteitä.

**Taulukko 5.6:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) lämpötilan suhteen vertailukohtana 434 Suomen alueella sijaitsevan Netatmo-aseman havainnot.

	$r^2$	RMSE (celsius)	$r^2$ (ääriarvot)	RMSE (ääriarvot)
T (METNO)	0.961	1.030	0.960	1.170
T (MNWC)	0.949	1.343	0.950	1.513
T (LAPS Scand)	0.949	1.402	0.950	1.605
T (LAPS Finland)	0.947	1.414	0.947	1.592



**Kuva 5.12:** Sama kuin kuvassa 5.11 mutta suhteelliselle kosteudelle.

**Taulukko 5.7:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) suhteellisen kosteuden suhteen vertailukohtana 494 Suomen alueella sijaitsevan Netatmo-aseman havainnot.

	$r^2$	RMSE (%)	$r^2$ (ääriarvot)	RMSE (ääriarvot)
RH (METNO)	0.560	9.933	0.723	9.453
RH (MNWC)	0.705	8.605	0.808	8.954
RH (LAPS Scand)	0.739	8.141	0.823	8.221
RH (LAPS Finland)	0.696	8.600	0.800	8.765



## 6. Pohdinta

Tässä tutkielmassa pohdintaosuus on jaettu kolmeen osaan. Aluksi luvussa 6.1 käydään läpi, millaisia johtopäätöksiä on tehty tutkimustuloksista niin että alaluvussa 6.1.1 esitellään johtopäätökset suure kerrallaan sekä käsitellään Netatmo-asemien tuomaa lisäarvoa. Alaluvussa 6.1.2 tarkastellaan maantieteellisten ongelma-alueiden syitä ja käydään läpi, millaisia johtopäätöksiä voidaan tehdä analyysien ääriarvo-, kuukausi- ja vuorokaudenaikajakaumista. Luvussa 6.2 vertaillaan tämän tutkimuksen tuloksia aiempaan tutkimustietoon ja lopuksi luvussa 6.3 arvioidaan tutkielman vahvuuksia ja heikkouksia sekä esitetään joitain jatkoehdotuksia.

### 6.1 Johtopäätökset data-analyysista

#### 6.1.1 Kokonaiskuva suureiden osalta ja Netatmo-asemien tuoma lisäarvo

Tarkastelemalla analysoituja pintasuureita sekä riippuvaa että riippumatonta havaintoaineistoa vasten voidaan muodostaa johtopäätöksiä ainakin järjestelmien kyvykkyydestä eri suureiden välillä. Sekä SYNOP-havaintoihin että Netatmo-havaintoihin perustuvaan vertailuun pohjautuen voidaan sanoa analyysien toistavan hyvin havaitun lämpötilan vaihtelun (taulukot 5.1 ja 5.6). Lisäksi etenkin merenpintapaineen analyysit olivat tarkkoja SYNOP-asemavertailun perusteella; kuten kuvasta 5.1b nähdään, olivat tilastolliset virheet merenpintapaineelle suurimmillaankin  $\pm 1,5$  hPa suurusluokkaa. Merenpintapaineanalyysien on todettu aiemminkin vastaavan hyvin havaintoja, sillä esimerkiksi King (2003) havaitsi tutkimuksessaan ECMWF-mallin merenpintapaineanalyysin toistavan tarkasti havaitun painevaihtelun Antarktiksien edustalla, jossa ylipäättään on hyvin vähän havaintotietoa saatavana. Suureista paineen vaihtelut ovat ajallisesti ja paikallisesti tasaisia, mikä realisoituu erinomaisina verifointituloksina.

Kosteus (ilmaistuna sekä RH:n että kastepisteen avulla) oli myös suurelta osin kuvattu hyvin analyyseissa etenkin kun suhteelliset kosteudet olivat alhaisia (kuva 5.6). Kosteuden tapauksessa METNO-analyysi selitti muita järjestelmiä heikommin havaitun kosteuden vaihtelun (pätee myös riippumattomalle tarkastelulle, kuva 5.12), mutta kuten luvussa 3.2.2 todettiin, muut kuin lämpötila-analyysit ovat METNO-järjestelmässä

tällä hetkellä MEPS-mallin vallitsevan ajanhetken ennusteita. Tämä voi selittää eroa, sillä MEPS päivittyy kolmen tunnin välein, jolloin havainnot eivät päivitty METNO-analyysihin samassa tahdissa muiden analyysien kanssa. Kosteus on parametrina altis isoillekin muutoksille etäisyyksien kasvaessa (Hubbard, 1994), joten verifiointin tulokset ovat rohkaisevia kosteusanalyysien kyvykkyyden suhteen. Verratessaan LAPS-analyysia riippumattomiin havaintoihin olivat Hiemstra et al. (2006) tulokset samansuuntaisia kuin tässä tutkimuksessa, ja he päättelivät tämän johtuvan onnistuneista lämpötilan ja kastepisteen analyyseistä. Koska mallien kastepistetulkintoja ei ollut saatavissa tätä tutkielmaa varten, voivat saadut hyvät tulokset kastepisteelle (kuva 5.8, taulukko A.4) tukea tätä päättelyketjua.

Tuulen nopeudessa on paljon alueellista vaihtelua (Hubbard, 1994), ja esimerkiksi pinnan rosoisuuden aiheuttama turbulenssi, maastonmuotojen aiheuttama kanavoituminen tai maastonkorkeus voivat vaikeuttaa tuulen ennustettavuutta (Hiemstra et al., 2006; Sharp et al., 2015). Tässä tutkimuksessa tuulen nopeuden verifiointitulokset olivatkin esimerkiksi lämpötilaa ja merenpintapainetta vaihtelevampia (kuva 5.9). Edellämainitut tekijät, kuten kanavoituminen ja turbulenssi, todennäköisesti vaikuttivat tähän, ja lisäksi on muistettava tuulen nopeuden olevan melko altis parametrisoituvirheille. Tarkasteltaessa suurimpia pistekohtaisia RMSE-arvoja (kuva 5.10) olivat keskivirheet suurimmillaan Lapissa, jossa on esimerkiksi tuntureista johtuen muuta maata suurempia pinnanvaihteluja. LAPS-analyysien virheet olivat pienempiä kahden muuhun analyysin verrattuna (kuva 5.1d), ja LAPS-analyysit myös korreloivat parhaiten havaitun vaihtelun kanssa (taulukko 5.5). Kuvasta 5.10 nähdään myös, että esimerkiksi korkeavirheisillä Lapin asemilla LAPS-analyysin keskivirheet olivat muita analyysejä pienempiä; tuulen nopeuden kohdalla voi siis sanoa olevan jo viitteitä mallien keskenäisestä paremmuudesta ottaen huomioon sen, ettei tässä tutkimuksessa ole tutkittu tilastollisia eroavaisuuksia aineistojen välillä. Kuten kosteuden tapauksessa, tuulen onnistunein analyysi LAPS-järjestelmissä voi olla seurausta esimerkiksi havaintojen vilkkaimmasta päivitysnopeudesta LAPS:ssa.

Analyysoijajärjestelmien (ja etenkin käytettyjen data-assimilaatiomenetelmien) keskenäisestä vertailusta on hankala tehdä johtopäätöksiä ilman kunnollista riippumattonta havaintoaineistoa; Netatmo-asemavertailuakin pystyttiin tekemään vain kahdelle parametrille, joista lämpötilan osalta havainnot eivät edes ole kaikille järjestelmille riippumattomia. Jos tätä kerättyä aineistoa voitaisiin verrata kaikki parametrit kattavaan laatu- ja tarkastettuun havaintoaineistoon, antaisi tämä paljon lisäarvoa tutkimukselle. Analyysoijien keskenäisen tarkoituksen on kuitenkin ratkaista parametrien jakauma myös niillä alueilla, joilla pistehavainnot ei ole käytettävissä. Tämän tutkimuksen alkuperäisenä suunnitelmiana oli verrata malliaineistoa neljän ns. SMEAR-tutkimusaseman havaintoaineistoon mutta tutkimuksen edetessä kävi ilmi, että muutaman kilometrin säteellä

jokaisesta SMEAR-asemasta on SYNOP-asema. SMEAR-asetat olisivat siis tuoneet hyvin vähän lisäarvoa tutkimukseen. Huomioitavaa on, että osa vertailuun päätyneistä Netatmo-asetamista sijaitsee myös SYNOP-asemien läheisyydessä mutta suuri osa myös ei.

Tämän työn puitteissa tieteellisesti kiinnostava motivaattori oli selvittää, vaikuttaako Netatmo-havaintojen käytöstä olevan analyysissä lisäarvoa. Tähän lähdettiin sillä hypoteesilla, että lisäarvoa olisi nähtävissä eli että tässä tapauksessa norjalaisen lämpötila-analyysin verifointitulokset olisivat lämpötila-analyysien joukosta parhaimmat. Näin tämän työn osalta kävikin, mutta on useastakin syystä hankala sanoa, ovatko juuri Netatmo-asetat aiheuttaneet havaitun lisäarvon vai ei. Kuten taulukossa 3.1 esitetään, on METNO:n lämpötila-analyysi selkeästi muita malleja hienohilaisempi, mikä voi olla tärkeä selittävä tekijä METNO-analyysin paremmuudessa (taulukko 5.1). Jos taas käydään läpi analyyseissä käytettyjä assimilaatiomenetelmiä (luku 2), on METNO:n lämpötila-analyysissä käytetty paikallista parven Kalman-muunnossuodatinta sekä parven optimaalista interpolointia yhdistävä assimilaatiomenetelmä muita käytettyjä menetelmiä kehittyneempi, mikä voi myös selittää analyysien kyvykkyyseroja.

Myös Netatmo-asemiin pohjautuvassa vertailussa METNO-analyysin verifointitulokset olivat analyyseistä parhaimmat (taulukko 5.6 sekä kuva A.3), mikä oli tietysti odotettavissa, koska METNO:n kohdalla kyse oli riippuvasta havaintovertailusta. METNO-analyysissä lämpötilahavaintojen järjestykseen tarkastelu on kuitenkin paljon tarkempaa kuin tässä tutkimuksessa, jolloin todennäköisesti havaintovertailussa on ollut mukana monia METNO-analyysillekin ns. riippumattomia havaintoja, joiden vaikutus METNO-analyysin verifointituloksiin on voinut olla jopa heikentävä. Silti virheet olivat METNO-analyysissä muita pienemmät (kuva A.3); toki tätä eroa voi edelleen selittää esimerkiksi samat syyt kuin esitettiin SYNOP-vertailun kohdalla, mutta ainakaan tämän tutkimuksen perusteella ei voi sanoa kansalaishavaintojen käytön heikentävän analyysin laatua.

### 6.1.2 Maantieteelliset ongelma-alueet ja lisätarkastelut

Yksi tämän tutkielman tavoitteista oli selvittää ilmenisikö tuloksittelun yhteydessä analyyseille ongelmallisia maantieteellisiä alueita. Kuten luvun 5.1 tulososiossa todettiin, oli etenkin lämpötilan, merenpintapaineen sekä tuulen nopeuden osalta huomattavissa suurimpien keskineliövirheiden painottuminen Lapin alueelle ja toisaalta LAPS-analyysin tapauksessa rannikkoalueiden ongelmallisuus. Tuulen nopeuden kohdalla on havaittu ainakin LAPS-analyysin tapauksessa keskivirheen kasvavan korkeuden funktiona (Hiemstra et al., 2006), mutta myös merenpintapaineelle voidaan olettaa samanlaista selitystä; tämä jo senkin vuoksi, että merenpintapaineiden laskemiseen on

käytetty järjestelmien korkeuksia. Tällöin alueet, joissa pinnanvaihteluja on enemmän, ovat alttiimpia eroille havainto- ja mallikorkeuksissa. Erot korkeuksissa taas voivat herkästi johtaa myös suurempiin keskivirheisiin näissä pisteissä - varsinkin jos toinen laskemiseen käytettävä suure eli lämpötila-analyysi on näillä alueilla myös virheellisempi kuin muualla, kuten on kyse Lapin tapauksessa; myös pohjoisen harvempi havaintoverkosto voi kasvattaa mallien virheitä.

Toinen huomio, mikä analyysien maantieteellisestä virhejakaumasta tehtiin oli, että LAPS-analyysin kohdalla merellisillä asemilla analyysivirheet olivat keskimäärin suurempia kuin maa-asemilla (merenpintapaineella ei kuitenkaan juuri eroa nähtävissä). Tätä voi selittää se, että verifioitavissa malleissa on taustalla erilaisia määritelmiä maa-meri-jakaumalle ja maankäytölle; esimerkiksi MEPS-säämallissa (ja täten sen nowcast-versiossa) käytetään maankäyttömäärittelyssä ECOCLIMAP-II-tietokantaa, jossa Euroopan alueeseen on kiinnitetty erityishuomiota (Müller et al., 2017). Analyysien topografia- ja maankäyttömäärittelysten erot voivat olla myös syy siihen, miksi LAPS:n lämpötila- ja kastepisteanalyyseissa on huomattavissa paikallista taipumusta yliennustamiseen (kuvat 5.2cd sekä 5.8cd). Tämä liittyy Lapin alueen analyysihin, sillä LAPS-analyysin keskivirheet olivat siellä analyyseista suurimpia (kuva 5.3), ja tätä yliarviointia oli havaittavissa vähemmän Netatmo-vertailussa (kuva 5.11), jossa ylipäättään Lapin asemat olivat aliedustettuina (kuva A.1). Todennäköisesti LAPS ei pysty aina ratkaisemaan ennustemalleillekin hankalia talvisia inversiotilanteita oikein.

Kokonaisten mittaussarjojen lisäksi tässä tutkielmassa otettiin tarkasteluun, miten hyvin analyysit onnistuivat toisintamaan suureiden havaitun vuorokausivaihtelun. On kuitenkin huomioitava, että todellisten vuorokausiminimien -ja maksimien analyyseja ei pystytty tutkimaan, sillä vertailua tehtiin tuntikohtaisia havaintokeskiarvoja vasten, joihin todelliset, hetkelliset minimi- ja maksimit ovat suodattuneet. Lisäksi on muistettava, että reilua vertailua kokonaisia mittaussarjoja vasten on vaikea suorittaa, koska ääriarvosarjojen otoskoko on paljon kokonaisia mittaussarjoja pienempi. Lämpötilan, merenpintapaineen ja kastepisteen vuorokausivaihtelut olivat kuitenkin kuvattu tuntikohtaisen vaihtelun tapaan hyvin analyyseissa (taulukot 5.1, 5.2, 5.4 ja 5.6), mikä ei ole yllättävää ottaen huomioon, että tutkittavista parametreista näiden suureiden vaihtelut ovat melko tasaisia ja helposti ennakoitavissa. Selkeämmät erot olivat huomattavissa tuulen nopeuden ja suhteellisen kosteuden tapauksissa, sillä niiden analyyseissa vuorokausivaihtelu oli toistettu tuntikohtaista vaihtelua paremmin (taulukot 5.3 ja 5.5). Todennäköisesti näiden suureiden analyyseissa on kuvattu paremmin vuorokausikohtainen kuin tuntikohtainen vaihtelu siitä syystä, että sekä suhteellinen kosteuden että tuulen nopeuden ajalliset muutokset ovat nopeita.

Liitteissä A.1 sekä A.2 esitetyt sekä kuukausikohtaiset (taulukot A.1-A.5) että vuo-

rokaudenaikakohtaiset (taulukko A.6) verifointitulokset täydentävät kokonaiskuvaa analyysien kyvykkyydestä Suomen alueella. Ensinnäkin merenpintapaineen osalta voidaan jo todeta, että kuukaudella tai vuorokaudenajalla ei ollut tässä tutkimuksessa suurtakaan merkitystä analyysin onnistumisessa (taulukot A.2 ja A.6). Kuukausikohtaisella tasolla huomattavinta oli lämpötila- ja kosteusanalyysien kasvaneet keskivirheet huhtikuussa (taulukko A.6), mikä on voinut johtua säätilanteesta. Tämän vuoden huhtikuussa sää oli selkeästi tavanomaista lämpimämpää<sup>1</sup>, mikä yhdistettynä lumipeitteen vaikutukseen usein heikentää lämpötilan ja kosteuden ennustettavuutta. Tällöin taustakentät ovat voineet olla tavallista virheellisempiä, mikä on voinut heijastua myös analyysien onnistumiseen. Ylipäättään kosteusanalyysit onnistuivat paremmin kesäisempinä kuukausina, mitä tukee myös hajontakuvioiden muodot eli pienempi hajonta alhaisilla RH-arvoilla (kuva 5.6); lämpimään ilmaan mahtuu enemmän kosteutta, jolloin suhteellinen kosteus on kesällä usein alhaisempi kuin talvella.

Analyysien kosteustulkinnoissa huomattiin enemmän ongelmia päivisin kuin öisin, kun tarkasteltiin keskivirheen suuruutta (taulukko A.3 ja taulukko A.4); lisäksi kosteusanalyyseista kyvykkäimmällä eli LAPS Scand-analyysillä kovin merkittävää päivä-yö-vaihtelua ei ollut havaittavissa. Suhteellisen kosteuden tapauksessa virheen pieneminen öisin selittyy sillä, että lämpötila- ja kastepisteanalyysit ovat keskimäärin öisin tarkempia (taulukko A.1 ja taulukko A.4), mutta tarkalleen sitä, miksi lämpötila- ja kastepisteanalyysit olivat virheellisempiä päivisin, on hankala sanoa; erot voivat olla osin tai lähes kokonaan kiinni sattumasta ja havaintojakson säätilanteista. Suhteellisen kosteuden kohdalla on kuitenkin todettava, että analyysit selittivät paremmin havaitun kosteuden vaihtelun päivisin kuin öisin (erot korrelaatiokertoimissa, taulukko A.3). Päivisin lämpötila on korkeimmillaan, jolloin suhteellinen kosteuskin on keskimäärin alhaisempi päivä- kuin yöaikaan. Kuten aiemmin todettiin, oli tässä työssä hajonta pienempää alhaisilla RH-arvoilla, mikä voi olla yksi korrelaatioeroa selittävä tekijä. Tuulen nopeuden suhteen verifointitulokset olivat paremmat öisin kuin päivisin (taulukko A.5). Päivisin tuuli on sekoittuneessa rajakerroksessa vaihtelevampaa kuin öisin, mikä voi näkyä kasvavana analyysivirheenä.

## 6.2 Vertailu aiempaan tutkimustietoon

Riippumattoman havaintoaineiston puuttuminen vaikeuttaa tulosten suhteuttamista aikaisempaan aiheesta tehtyyn tutkimukseen, sillä niissä analyysitietoa verrataan juurikin riippumattomiin aineistoihin. Hiemstra et al. (2006) tutkimuksessa havaittiin kuitenkin hyvin samankaltaisia yhteyksiä kuin tässä tutkimuksessa havaintoaineiston sekä

<sup>1</sup><http://www.ilmastokatsaus.fi/2019/05/24/huhtikuun-2019-kuukausikatsaus/>; sivulla käyty 11.9.2019

analyysin välillä. Lämpötilan osalta myös Hiemstra et al. (2006) tutkimuksessa lineaarinen regressioanalyysi tuotti lähelle 1:1-suoraa vastaavan sovituksen  $r^2$ -tunnusluvun ollessa 0.98. Keskivirhe on kuitenkin suurempi heidän tutkimuksessaan ( $1.8^\circ\text{C}$ ) kuin millään analyysijärjestelmällä tässä tutkimuksessa (laskien mukaan myös Netatmo-vertailun). Samansuuntainen huomio pätee myös kosteudelle sekä tuulelle, ja tämä ero keskivirheissä voi selittyä esimerkiksi eroilla aineistojen laadussa, itse järjestelmissä (tutkimus tehty vuonna 2006) sekä otannan laajuudessa; Hiemstra et al. (2006) tutkimuksessa oli mukana Suomen aluetta suurempi alue suurempine pinnanvaihteluineen.  $r^2$ -luvut LAPS-analyysin kosteudelle ja tuulelle ovat Hiemstra et al. (2006) tutkimuksessa taas tässä järjestyksessä 0.87 ja 0.52; tuulen nopeuden selkeästi heikompi korrelaatio voi esimerkiksi johtua juurikin eroista riippumattoman aineiston ja SYNOP-asemien välillä.

Vitali et al. (2008) tutkimuksessa heidän havaintonsa lämpötilan ja kosteuden analyysista ovat yhteneväisiä tämän tutkimuksen ja Hiemstra et al. (2006) tulosten kanssa. Kahdesta suureesta lämpötilan osalta tulokset ovat Vitali et al. (2008) tutkimuksessa paremmat;  $r^2$ -luvut ovat n. 0.95 suuruusluokkaa, RMSE-arvot  $1\text{--}2^\circ\text{C}$ , ja lineaarinen sovitus vastaa myös heidän raportissaan hyvin 1:1-suoraa. Tuulianalyysin hajontakuviossa on selkeästi lämpötilaa enemmän vaihtelua Vitali et al. (2008) raportissa, mikä heijastuu myös tunnuslukuihin;  $r^2$ -luvut ovat 0-0.4 välillä ja RMSE-arvot yli  $2\text{ m/s}$ . Vitali et al. (2008) verifioivat tuloksia vain kahden vertailupisteen osalta, eli otanta on esimerkiksi Hiemstra et al. (2006) verrattuna paljon pienempi. LAPS-analyysija on verifioitu myös Farguell et al. (2016) tutkimuksessa, jossa verrattiin kolmea eri analyysimallia (yhtenä LAPS) lämpötilan, kosteuden ja tuulen nopeuden osalta 77 riippumattoman havaintoaseman tietoihin vuonna 2012 Katalonian alueella. RMSE-arvojen sijaan Farguell et al. (2016) käyttivät raportissaan tunnuslukuna MAE-arvoa (Mean Absolute Error), joka on vertailukelpoinen RMSE-arvoihin. LAPS-analyysille he havaitsivat n.  $1.7^\circ\text{C}$ ,  $8.1\%$  sekä  $1.3\text{ m/s}$  RMSE-arvoja vastaavat virheet, jotka tukevat niin tämän tutkimuksen kuin muiden vastaavien tutkimusten havaintoja.

## 6.3 Tutkimuksen arviointi ja jatkoehdotukset

Tämän tutkielman suurin ansio on se, että on onnistuttu luomaan kokonaiskatsaus siihen, miten analyysijärjestelmät kokonaisuudessaan onnistuvat parametreittain Suomen alueella. Tutkimusasetelmien erot huomioon ottaen tässä työssä saadut tulokset ovat linjassa aikaisempien tutkimusten kanssa ja rohkaisevat erityisesti lämpötilan, paineen sekä kosteudenkin analyysien käyttöön. Tämä on arvokasta tietoa käyttäjän eli Ilmatieteen laitoksen näkökulmasta, sillä tulevaisuudessa säänennusteiden tuotannossa automaatio on lisääntymässä ja niin lähituntien kuin kymmenen vuorokauden ennuste-

den tuotantoprosesseja kehitetään. Käyttäjän näkökulmasta on hyödyllistä myös tietää taustakentältään MEPS-pohjaiseen LAPS-analyysiin siirtymisen tuoneen lisäarvoa ainakin kosteusanalyysiin. Lisäksi data-assimilaatio on aiheena hyvin tärkeässä osassa meteorologian alalla, ja tämän tutkielman kirjallisuuskatsauksena toimiva teoriaosuus (luku 2) täydentää hyvin aiempia aihetta käsitteleviä tai sivuavia muita suomenkielisiä pro gradu-tutkielmia (esim. Iipponen, 2017; Alatalo, 2009; Hakkarainen, 2009).

Riippumattoman havaintoaineiston puuttuminen ja aineiston suuri koko rajoittivat analyysien keskinäisen vertailun tekemistä. Data-analyysi oli myös tämän työn vaatimissa puitteissa rajattava melko yksinkertaiseksi, eli verifiointissa keskityttiin kokonaisten malli- ja havaintosarjojen lineaariseen regressioanalyysiin, eikä aineistojen tilastollisiin eroihin otettu kantaa. Olisi ollut mielenkiintoista tutkia myös tapaustutkimuksena yksittäisten asemien aikasarjoja esimerkiksi hankalalta Lapin alueelta, sillä tämä olisi voinut tukea tätä pohdintaosuutta. Lisäksi merenpintapaineen ja kastepisteen verifiointitulokset riippuivat tässä tutkimuksessa tietysti käytetyistä kaavoista (luku 4.2); tapoja laskea kastepiste tai merenpintapaine on useita, ja niiden vertailu olisi voinut tuoda lisäarvoa tutkimukselle. Analyysien uudelleenlaskentaan olisi taas voitu käyttää lähimmän naapurin menetelmää hienostuneempia menetelmiä, mutta lähimmän naapurin menetelmän havaittiin toimivan tarpeeksi hyvin tämän tutkimuksen tarpeisiin nähden.

Jos tätä tutkimusta haluaisi tästä mittakaavasta laajentaa, voisi olla hyödyllistä tutkia esimerkiksi analyysien keskivirheiden sekä korkeuden välistä riippuvuutta, ja etenkin tuulen nopeuden osalta tuloksia voisi verrata Hiemstra et al. (2006) tutkimukseen. Yksi vaihtoehto olisi myös kerätä uutta analyysiaineistoa niin että analyysiin syötettävistä SYNOP-havainnoista osa jätettäisiin aineiston keruun ajaksi syöttämättä, jolloin data-analyysin vastejoukkona olisi oikeasti riippumattomia sekä laatuvarmistettuja havaintoja. Tämän tutkimuksen riippumatonta asemavertailua vastaavan verifiointin tapauksessa on liitteissä mukana kuvia 5.3 sekä 5.7 vastaavat kuvat Netatmo-asemavertailulle (kuva A.1 sekä A.2), mutta niistä ei ole kovin suurta lisäarvoa ajatellen analyysien keskivirheen suuruutta johtuen hyvin yksinkertaisesta suoritetusta laatutarkastelusta - todennäköisesti suurimmat keskivirheet johtuvat pikemmin itse havaintoaineiston virheellisyydestä. Netatmo-asemien tehokas hyödyntäminen tässä tutkimuksessa olisi kuitenkin vaatinut tutkielman laajuuden kannalta liikaa työtä ottaen huomioon jo olemassaolevan analysoitavan aineiston määrän. Suunnitelmissa on, että LAPS-järjestelmään aletaan Ilmatieteen laitoksella syöttää Netatmo-kansalaishavaintoja, minkä pohjalta voi tulevaisuudessa lähteä tekemään uutta verifiointia, jos aineiston laatutarkastelusta pidetään hyvää huolta.

## 7. Yhteenveto

Ilmatieteen laitoksella lähihetken ennusteita korjataan hyödyntämällä data-assimilaation avulla tuotettuja pinta-analyyseja. Tässä tutkielmassa verifioitiin kolmen Ilmatieteen laitoksella käytettävän analyysi- ja nowcast-mallin pinta-analyyseja sekä SYNOP-että Netatmo-havaintoverkostoja vasten. Tarkoituksena oli luoda kokonaiskuva analyysien onnistumisesta lämpötilan, paineen, kosteuden ja tuulen analysoinnissa Suomen alueella lineaarisen regressioanalyysin avulla. Erityisesti lämpötilan ja merenpintapaineen analyysien havaittiin olevan Suomen alueella laadukkaita; lämpötila-analyysien keskivirhe oli 0.1-0.7°C ja merenpintapaineen 0.4 hPa tienoilla SYNOP-asemavertailun mukaan. Lämpötilasta oli lisäksi käytettävissä riittävän laadukkaita Netatmo-havaintoja mielekästä verifiointia varten: näin epävarmempiakin havaintosarjoja vasten lämpötila-analyysien keskivirheen havaittiin vaihtelevan 1.0-1.3°C välillä, mikä on linjassa aikaisempien tutkimusten kanssa huomioiden tutkimusasetelmien erot. Kosteus- ja tuulianalyysien kohdalla hajontaa oli lämpötilaa ja painetta enemmän, mutta kosteudelle sekä tuulelle on ominaista voimakas alueellinen ja ajallinen vaihtelu, mikä selittää analyysien laatueroja.

Tutkimalla yksittäisten havaintopisteiden RMSE-arvoja pystyttiin selvittämään, millä alueilla suurimmat analyysivirheet ilmenevät. Eniten ongelmia havaittiin merellisillä asemilla sekä Lapissa. Rannikkoalueilla ongelmia oli eniten LAPS-analyyseilla, mikä todennäköisesti johtui muista malleista eroavasta maa-merijakauman kuvauksesta; Lapin alueella analyysien ongelmat liittyivät taas topografian kuvausongelmiin. Analyysimallien resoluutio ei aina riitä ratkaisemaan Lapin pinnanvaihteluita, ja esimerkiksi lämpötila-analyyseista hienohilaisin METNO-analyysi onnistui siellä parhaiten.

Norjan ilmatieteen laitoksen kehittämä analyysijärjestelmä valikoitui vertailuun varsinkin siitä syystä, että heidän lämpötila-analyysissan hyödynnetään kansalaishavaintoja Netatmo-asemien toimittaman datan muodossa. Yksi alkuperäisistä tutkimustarkoituksista oli selvittää, parantaako kansalaishavaintojen käyttö analyysin laatua. Norjalainen METNO-analyysi havaittiin lämpötilan osalta tarkimmaksi, mutta varsinkin ilman riippumattomia havaintoja ei kysymykseen pystytty vastaamaan, sillä METNO-analyysin paremmuus SYNOP-asemavertailussa voi hyvinkin selittyä eroilla resoluutioissa, assimilaatiomenetelmissä tai syötetyissä havaintoaineistoissa. Jos ana-



lyysijärjestelmien suhteen riippumatonta havaintoaineistoa olisi mahdollista saada jos-sain kohtaa tulevaisuudessa käyttöön, voitaisiin tätä kautta analyysijärjestelmiä verrata paremmin keskenään ja esimerkiksi viemällä analyysit yhteiseen hilaan vastata kysymykseen kansalaishavaintojen hyödystä. Nykyisessä muodossa tämän tutkimuksen tuloksista tulee kuitenkin olemaan jo apua lähituntien ennusteiden tuotantoprosessien parantamisessa Ilmatieteen laitoksella.

# Kirjallisuutta

- M. Alatalo. Havaintojärjestelmäkokeet numeerisessa sään ennustamisessa. 2009. Pro Gradu.
- S. C. Albers, J. A. McGinley, D. L. Birkenheuer, ja J. R. Smart. The Local Analysis and Prediction System (LAPS): Analyses of clouds, precipitation, and temperature. *Weather and Forecasting*, 11(3):273–287, 1996.
- U. Andrae. The MetCoOp ensemble MEPS. *ALADIN-HIRLAM Newsletter*, (8):98–104, 2017.
- S. L. Barnes. A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, 3(4):396–409, 1964.
- L. Bengtsson, U. Andrae, T. Aspelien, Y. Batrak, J. Calvo, W. de Rooy, E. Gleeson, B. Hansen-Sass, M. Homleid, M. Hortal, et al. The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system. *Monthly Weather Review*, 145(5):1919–1935, 2017.
- P. Bergthörsson ja B. R. Döös. Numerical weather map analysis. *Tellus*, 7(3):329–340, 1955.
- D. Birkenheuer. The effect of using digital satellite imagery in the LAPS moisture analysis. *Weather and forecasting*, 14(5):782–788, 1999.
- A. Buck. Buck research cr-1a user’s manual. *Buck Research Instruments: Boulder, CO, USA*, 1996.
- L. Chapman, C. Bell, ja S. Bell. Can the crowdsourcing data paradigm take atmospheric science to a new level? a case study of the urban heat island of London quantified using Netatmo weather stations. *International Journal of Climatology*, 37(9):3597–3605, 2017.
- G. P. Cressman. An operational objective analysis system. *Mon. Wea. Rev.*, 87(10):367–374, 1959.
- R. Daley. *Atmospheric data analysis*. Number 2. Cambridge university press, 1993.

- S. Dance, E. Ebert, ja D. Scurrah. Thunderstorm strike probability nowcasting. *Journal of Atmospheric and Oceanic Technology*, 27(1):79–93, 2010.
- A. Eliassen. Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. *Peport no 5, Videnskaps-Akademiets Institutt for Vaer-Og Klimaforskning, Oslo, Norway*, 12, 1954.
- G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.
- G. Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367, 2003.
- G. Evensen. Sampling strategies and square root analysis schemes for the EnKF. *Ocean dynamics*, 54(6):539–560, 2004.
- A. Farguell, J. Moré, A. Cortés, J. R. Miró, T. Margalef, ja V. Altava. Reducing Data Uncertainty in Surface Meteorology using Data Assimilation: A Comparison Study. *Procedia Computer Science*, 80:1846–1855, 2016.
- L. Gandin. S, 1963: Objective analysis of meteorological fields. *Translated from Russian by the Israeli Program for Scientific Translations*, 1965.
- E. Gregow. *New methods using in-situ and remote-sensing observations for improved meteorological analysis*. Väitöskirja, Helsingin yliopisto, Suomi, 2018. ISBN:978-952-336-042-6.
- L. Häggmark, K.-I. Ivarsson, S. Gollvik, ja P.-O. Olofsson. Mesan, an operational mesoscale analysis system. *Tellus A: Dynamic Meteorology and Oceanography*, 52(1):2–20, 2000.
- T. Haiden, A. Kann, C. Wittmann, G. Pistotnik, B. Bica, ja C. Gruber. The Integrated Nowcasting through Comprehensive Analysis (INCA) system and its validation over the Eastern Alpine region. *Weather and Forecasting*, 26(2):166–183, 2011.
- J. Hakkarainen. Data-assimilaatiomenetelmistä ja niiden soveltamisesta ROSE-kemiakuljetusmalliin. 2009. Pro Gradu.
- C. A. Hiemstra, G. E. Liston, R. A. Pielke Sr, D. L. Birkenheuer, ja S. C. Albers. Comparing Local Analysis and Prediction System (LAPS) assimilations with independent observations. *Weather and forecasting*, 21(6):1024–1040, 2006.
- P. L. Houtekamer ja H. L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3):796–811, 1998.

- K. Hubbard. Spatial variability of daily weather variables in the high plains of the USA. *Agricultural and Forest Meteorology*, 68(1-2):29–41, 1994.
- B. R. Hunt, E. J. Kostelich, ja I. Szunyogh. Efficient data assimilation for spatio-temporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1-2):112–126, 2007.
- J. Iipponen. Data-assimilaatio puoliempiirissä termosfäärimallissa. 2017. Pro Gradu.
- T. Janjić, N. Bormann, M. Bocquet, J. Carton, S. Cohn, S. Dance, S. Losa, N. Nichols, R. Potthast, J. Waller, et al. On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1257–1278, 2018.
- K. Jasper, J. Gurtz, ja H. Lang. Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. *Journal of hydrology*, 267(1-2):40–52, 2002.
- R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.
- E. Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge university press, 2003.
- J. King. Validation of ECMWF sea level pressure analyses over the Bellingshausen Sea, Antarctica. *Weather and forecasting*, 18(3):536–540, 2003.
- J. T. Koskinen, J. Poutiainen, D. M. Schultz, S. Joffre, J. Koistinen, E. Saltikoff, E. Gregow, H. Turtiainen, W. F. Dabberdt, J. Damski, et al. The Helsinki Testbed: a mesoscale measurement, research, and service platform. *Bulletin of the American Meteorological Society*, 92(3):325–342, 2011.
- W. Lahoz. Research satellites. Teoksessa K. B. Lahoz, William A ja R. Ménard, toimittajat, *Data Assimilation*, sivut 301–321. Springer, 2010.
- W. A. Lahoz ja P. Schneider. Data assimilation: making sense of Earth Observation. *Frontiers in Environmental Science*, 2:16, 2014.
- M. G. Lawrence. The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bulletin of the American Meteorological Society*, 86(2):225–234, 2005.
- G. E. Liston ja K. Elder. A distributed snow-evolution modeling system (SnowModel). *Journal of Hydrometeorology*, 7(6):1259–1276, 2006.

- E. N. Lorenz. A study of the predictability of a 28-variable atmospheric model. *Tellus*, 17(3):321–333, 1965.
- J. A. McGinley, S. C. Albers, ja P. A. Stamus. Validation of a composite convective index as defined by a real-time local analysis system. *Weather and Forecasting*, 6(3):337–356, 1991.
- H. L. Mitchell, P. L. Houtekamer, ja G. Pellerin. Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Monthly weather review*, 130(11):2791–2808, 2002.
- M. Müller, M. Homleid, K.-I. Ivarsson, M. A. Køltzow, M. Lindskog, K. H. Midtbø, U. Andrae, T. Aspelien, L. Berggren, D. Bjørge, et al. AROME-MetCoOp: A Nordic convective-scale operational weather prediction model. *Weather and Forecasting*, 32(2):609–627, 2017.
- National Research Council. *Observing weather and climate from the ground up: A nationwide network of networks*. National Academies Press, 2009.
- T. Nipen, I. Seiderstad, C. Lussana, N. E. Larsgård, ja M. Wolff. Crowdsourced data improves temperature forecasts on Yr.no, 2015. URL [https://library.wmo.int/doc\\_num.php?explnum\\_id=5377](https://library.wmo.int/doc_num.php?explnum_id=5377). Sivustolla käyty 16.11.2019.
- P. R. Oke, G. B. Brassington, D. A. Griffin, ja A. Schiller. Ocean data assimilation: a case for ensemble optimal interpolation. *Australian Meteorological and Oceanographic Journal*, 59(Sp. Iss):67–76, 2010.
- E. Ruzanski, V. Chandrasekar, ja Y. Wang. The CASA nowcasting system. *Journal of Atmospheric and Oceanic Technology*, 28(5):640–655, 2011.
- I. A. Seierstad. Henkilökohtainen tiedonanto. 01.02.2019.
- E. Sharp, P. Dodds, M. Barrett, ja C. Spataru. Evaluating the accuracy of CFSR reanalysis hourly wind speed forecasts for the UK, using in situ measurements and geographical information. *Renewable Energy*, 77:527–538, 2015.
- J. S. Snook, J. M. Cram, ja J. M. Schmidt. LAPS/RAMS. a nonhydrostatic mesoscale numerical modeling system configured for operational use. *Tellus A*, 47(5):864–875, 1995.
- I. Szunyogh, E. A. Satterfield, J. A. Aravéquia, E. J. Fertig, G. Gyarmati, E. Kalnay, B. R. Hunt, E. J. Kostelich, D. D. Kuhl, E. Ott, et al. The local ensemble transform Kalman Filter and its implementation on the NCEP global model at the University of Maryland. Teoksessa *Proceedings of the ECMWF Workshop on Flow dependent aspects of data assimilation*, sivut 47–64, 2007.

- O. Talagrand. Assimilation of Observations, an Introduction. *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):191–209, 1997.
- O. Vignes ja U. Andrae. MetCoOp status and plans. Poster esitetty tapaamisessa ”Joint 29th ALADIN Workshop and HIRLAM All Staff Meeting 2019”, 1-5 huhtikuuta, Madrid, Espanja, 2019.
- L. Vitali, G. Pace, G. Zanini, G. Bracco, S. Migliori, ja S. Finardi. Application of the Local Analysis and Prediction System (LAPS) to carry out high resolution meteorological analysis at Italian national scale: implementation and first results for year 1999. Technical report, Citeseer, 2008.
- J. W. Wilson, E. E. Ebert, T. R. Saxen, R. D. Roberts, C. K. Mueller, M. Sleight, C. E. Pierce, ja A. Seed. Sydney 2000 forecast demonstration project: convective storm nowcasting. *Weather and forecasting*, 19(1):131–150, 2004.
- World Meteorological Organization. Pressure Reduction Formula. Dokumentti esitetty työryhmän ”CIMO Expert Team on Standardization” 1. sessiossa, 26-29 marraskuuta, Geneve, Sveitsi, 2012.

# A. Data-analyysin lisätuloksia

## A.1 Päivä- ja yökohtaiset tulokset

**Taulukko A.1:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) lämpötilan suhteen eriteltynä päivä- ja yöarvoihin vertailukohtana 197 suomalaisen pinta-sääaseman havainnot.

	$r^2$ (päivä)	RMSE (päivä)	$r^2$ (yö)	RMSE (yö)
T (METNO)	0.997	0.181	0.997	0.091
T (MNWC)	0.986	0.734	0.983	0.712
T (LAPS Scand)	0.985	0.658	0.974	0.677
T (LAPS Finland)	0.982	0.668	0.978	0.628

**Taulukko A.2:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) merenpintapaineen suhteen eriteltynä päivä- ja yöarvoihin vertailukohtana 197 suomalaisen pinta-sääaseman havainnot.

	$r^2$ (päivä)	RMSE (päivä)	$r^2$ (yö)	RMSE (yö)
P (METNO)	0.998	0.410	0.998	0.389
P (MNWC)	0.998	0.372	0.998	0.380
P (LAPS Scand)	0.998	0.473	0.998	0.412
P (LAPS Finland)	0.997	0.444	0.997	0.456

**Taulukko A.3:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) suhteellisen kosteuden suhteen eriteltynä päivä- ja yöarvoihin vertailukohtana 197 suomalaisen pinta-sääaseman havainnot.

	$r^2$ (päivä)	RMSE (päivä)	$r^2$ (yö)	RMSE (yö)
RH (METNO)	0.666	9.618	0.464	7.822
RH (MNWC)	0.901	4.357	0.829	4.046
RH (LAPS Scand)	0.919	3.606	0.844	3.419
RH (LAPS Finland)	0.877	4.295	0.796	3.697

**Taulukko A.4:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) kastepisteen suhteen eriteltynä päivä- ja yöarvoihin vertailukohtana 197 suomalaisen pinta-sääaseman havainnot.

	$r^2$ (päivä)	RMSE (päivä)	$r^2$ (yö)	RMSE (yö)
$T_d$ (METNO)	0.871	2.213	0.937	1.428
$T_d$ (MNWC)	0.974	0.811	0.981	0.711
$T_d$ (LAPS Scand)	0.983	0.621	0.977	0.585
$T_d$ (LAPS Finland)	0.965	0.834	0.975	0.641

**Taulukko A.5:** Eri analyysijärjestelmien korrelaatiokertoimet ja keskineliövirheen neliöjuuret (RMSE) tuulen nopeuden suhteen eriteltynä päivä- ja yöarvoihin vertailukohtana 197 suomalaisen pinta-sääaseman havainnot.

	$r^2$ (päivä)	RMSE (päivä)	$r^2$ (yö)	RMSE (yö)
WS (METNO)	0.679	1.210	0.714	1.115
WS (MNWC)	0.668	1.225	0.694	1.134
WS (LAPS Scand)	0.791	0.913	0.806	0.882
WS (LAPS Finland)	0.800	0.912	0.807	0.876

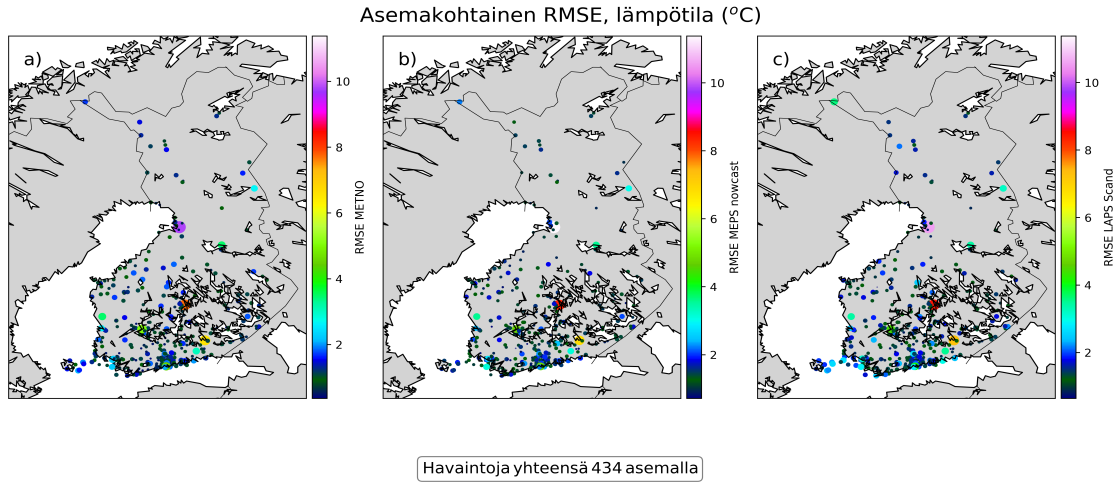


## A.2 Kuukausittaiset tulokset

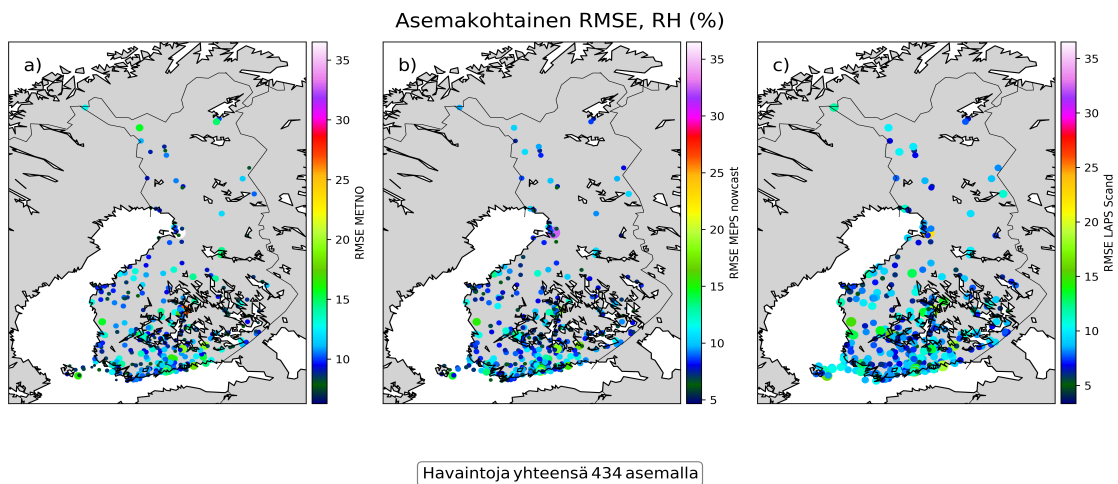
**Taulukko A.6:** Eri analyysijärjestelmien  $r^2$ -luvut ja keskineliövirheen neliöjuuret (RMSE) kuukausittain kaikille havaintosuureille vertailukohtana 197 suomalaisen pintasääaseman havainnot.

	maaliskuu	huhtikuu	toukokuu	kesäkuu
T(METNO)	0.995,0.141	0.988,0.184	0.995,0.076	0.991,0.149
T(MNWC)	0.971,0.658	0.950,0.863	0.973,0.650	0.962,0.723
T(LAPS Scand)	0.935,0.710	0.949,0.751	0.966,0.623	0.956,0.660
T(LAPS Finland)	0.953,0.641	0.940,0.768	0.964,0.588	0.953,0.644
P(METNO)	0.996,0.469	0.997,0.374	0.998,0.370	0.996,0.373
P(MNWC)	0.997,0.445	0.997,0.337	0.999,0.343	0.997,0.352
P(LAPS Scand)	0.997,0.527	0.997,0.402	0.998,0.426	0.996,0.409
P(LAPS Finland)	0.995,0.541	0.993,0.420	0.997,0.421	0.995,0.400
RH(METNO)	0.637,7.407	0.575,12.822	0.681,8.840	0.727,7.472
RH(MNWC)	0.862,4.333	0.850,4.959	0.942,3.383	0.918,3.773
RH(LAPS Scand)	0.879,3.307	0.885,4.106	0.925,3.428	0.911,3.549
RH(LAPS Finland)	0.860,3.392	0.802,5.569	0.896,3.837	0.903,3.584
$T_d$ (METNO)	0.929,1.301	0.561,2.793	0.741,1.934	0.781,1.664
$T_d$ (MNWC)	0.962,0.808	0.858,0.894	0.970,0.618	0.958,0.682
$T_d$ (LAPS Scand)	0.934,0.711	0.929,0.590	0.971,0.560	0.959,0.614
$T_d$ (LAPS Finland)	0.949,0.705	0.816,0.705	0.942,0.707	0.951,0.645
WS(METNO)	0.704,1.262	0.726,1.09	0.705,1.127	0.654,1.155
WS(MNWC)	0.686,1.287	0.692,1.135	0.696,1.142	0.655,1.151
WS(LAPS Scand)	0.801,1.016	0.809,0.834	0.814,0.853	0.786,0.855
WS(LAPS Finland)	0.829,0.953	0.784,0.873	0.814,0.866	0.786,0.880

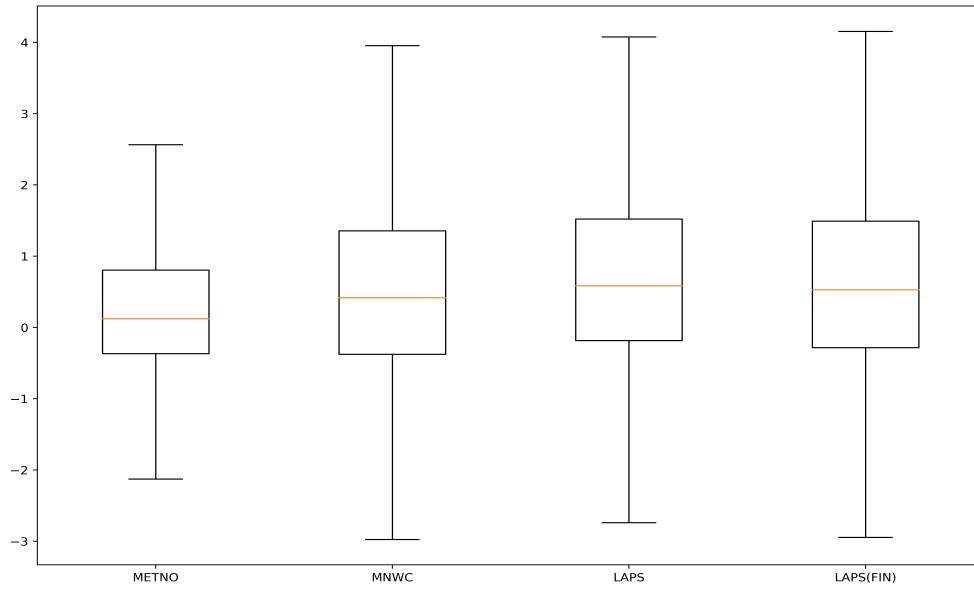
## A.3 Netatmo-vertailu



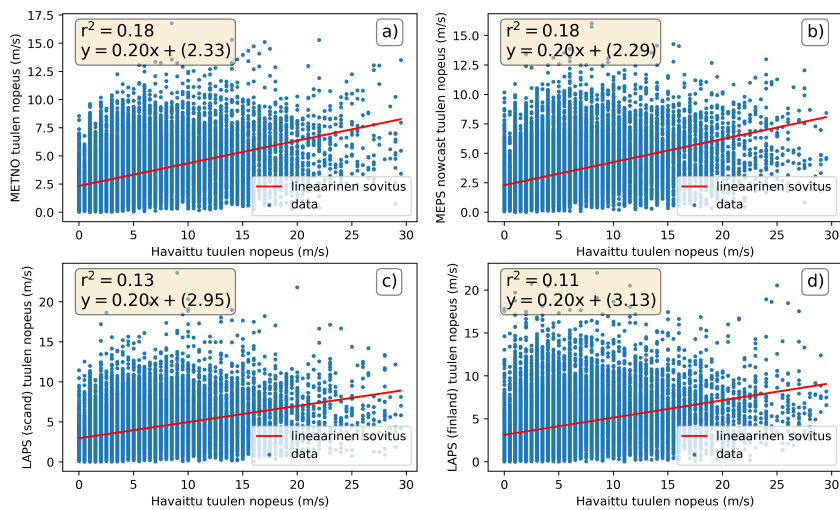
**Kuva A.1:** Analyysijärjestelmien laskemien suhteellisten kosteuksien keskineliövirheen neliöjuuret (RMSE) ilmoitettuna jokaiselle mittausjoukossa mukana olleelle lämpötilaa havaitsevalle Netatmo-sääasemalle erikseen. Kuvattuna on a) MET Nordic Analysis, b) MEPS-nowcast ja c) LAPS (Skandinavia). Lämpötilahavaintoja on kerätty yhteensä 434 asemalla Suomen alueella.



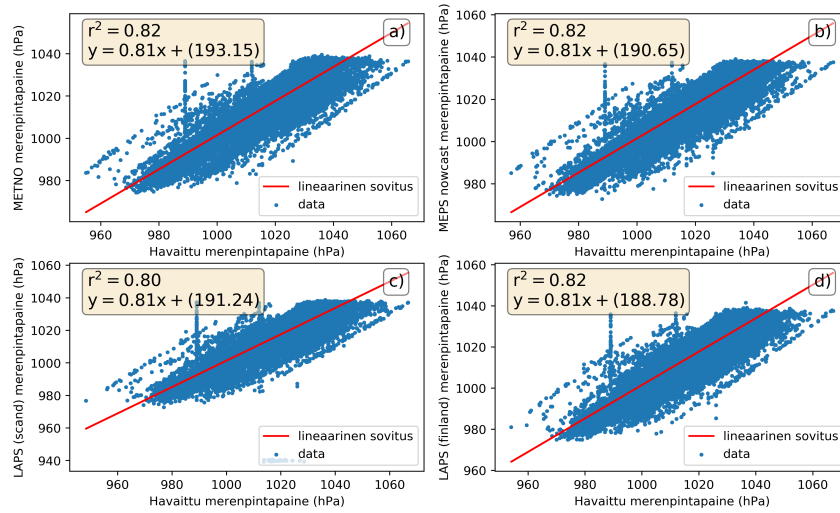
**Kuva A.2:** Sama kuin kuva A.1 mutta suhteelliselle kosteudelle.



**Kuva A.3:** Tilastolliset virheet esitettynä Netatmo-verailun lämpötilalle. Virheet (y-akseli) ovat esitetty ruutu- ja janakuviaina, joissa mediaani on esitetty punaisen viivan avulla, 25 ja 75-persentiilit ruutujen ala- ja ylärajojen avulla ja maksimit sekä minimit janojen avulla; x-akselilla on käytetty analyysijärjestelmät.



**Kuva A.4:** Sama kuin kuva 5.11 mutta tuulen nopeudelle.



**Kuva A.5:** Sama kuin kuva 5.11 mutta merenpintapaineelle.